

Geometry of Permutation Limits

Mustazee Rahman

Bálint Virág

Máté Vizer

Abstract

This paper initiates a limit theory of permutation valued processes, building on the recent theory of permutons. We apply this to study the asymptotic behaviour of random sorting networks. We prove that the Archimedean path, the conjectured limit of random sorting networks, is the unique path from the identity to the reverse permuton of minimal energy in an appropriate metric. This provides support for the Archimedean path conjecture. Together with a recent large deviations result (Kotowski, 2016) it also implies this conjecture for the model of relaxed random sorting networks.

1 Introduction

The objective of this paper is two-fold. First, it develops a limit theory of permutation valued processes with the idea that it will be applicable to study asymptotic properties of permutations that arise in combinatorics and probability theory. Second, it applies the limit theory to study the asymptotic behaviour of uniform random sorting networks.

1.1 Permutations limits

The asymptotic behaviour of permutations has been studied in combinatorics, probability theory and statistics. More recently, the language of permutons has been developed to study permutation limits. Examples within this theory include the study of finite forcibility [7], pattern avoidance [10], quasi-randomness [12], the Mallows model [18] and other models [4, 14]. In order to setup our limit theory we begin with a discussion of permutons.

Let \mathcal{S}_n denote the symmetric group of order n . For $\sigma \in \mathcal{S}_n$, its *empirical measure* is

$$\mu^\sigma = \frac{1}{n} \sum_i \delta_{\left(\frac{2i}{n}-1, \frac{2\sigma(i)}{n}-1\right)}. \quad (1.1)$$

This is a probability measure of $[-1, 1]^2$. One defines a sequence of permutations $\{\sigma_n\}$ with size $|\sigma_n| \rightarrow \infty$ to converge if μ^{σ_n} converges weakly to a Borel measure μ .

A *permuton* μ is a Borel probability measure on $[-1, 1]^2$ with uniform marginals. It is proven in [8] that limits of permutations in the above sense are precisely the permutons. Namely, limits of permutations are permutons and any permuton may be realized as a limit of permutations. The terminology ‘permuton’ is due to Glebov et. al. [7].

In this paper we study not a sequence of single permutations but rather a sequence of sequences

of permutations. For an integer n , let $[n] = \{1, \dots, n\}$. Suppose that

$$\sigma^n = (\sigma_t^n; t \in [t_n]) \quad (1.2)$$

is a \mathcal{S}_n -valued sequence. We say that σ^n is a *permutation process* of \mathcal{S}_n , or just a permutation process when there is no ambiguity. We always set σ_0^n to be the identity. Our first goal is to develop a limit theory of sequences of permutation processes growing both in the size, n , of the permutations and the length, t_n , of the sequence.

An important example of permutation processes that we wish to study are *sorting networks*. A sorting network of \mathcal{S}_n is a path of minimal length from the identity permutation $\mathbf{id}_n = 1, \dots, n$ to the reverse permutation $\mathbf{rev}_n = n, n-1, \dots, 1$ in the Cayley graph of \mathcal{S}_n generated by adjacent transpositions $\tau_i = (i, i+1)$ for $1 \leq i \leq n-1$. It is a permutation process with $t_n = \binom{n}{2}$, namely: $\mathbf{id}_n = \sigma_0, \dots, \sigma_{\binom{n}{2}} = \mathbf{rev}_n$, where $\sigma_k = \sigma_{k-1} \circ \tau_{i_k}$ for some τ_{i_k} . The adjacent transposition τ_i are also called swaps. An example of a sorting network is the classical bubble sort algorithm of computer science applied to \mathbf{rev}_n and viewed in reverse time.

Another example to which our limit theory applies is the interchange process. In this setting the resulting permutation process is random. This process turns out to converge in probability, and the limit is a deterministic object, the law of stationary Brownian motion. This is explained in Section 2.1.

Limits of permutation processes. Now we describe the setup of the limit theory and the first main result. Given a permutation process σ^n of \mathcal{S}_n as in (1.2), the rescaled trajectory of particle $i \in [n]$ is the function

$$T_i^n(t/t_n) = \frac{2\sigma_t^n(i)}{n} - 1 \text{ for } t \in [t_n]. \quad (1.3)$$

After linearly interpolating between the discrete times t/t_n , we may consider T_i^n as a continuous function from $[0, 1]$ to $[-1, 1]$. The *trajectory process* of σ^n , denoted X^n , is the trajectory of a particle chosen uniformly at random:

$$X^n = \frac{1}{n} \sum_{i=1}^n \delta_{T_i^n}. \quad (1.4)$$

Let \mathbf{C} denote the space of continuous functions from $[0, 1]$ to $[-1, 1]$ in the topology of uniform convergence. The trajectory process is then a Borel probability measure on \mathbf{C} .

Observe that for every $t \in [t_n]$ the distribution of $X^n(t/t_n)$ is uniform on the set $\{\frac{2i}{n} - 1; i \in [n]\}$ due to σ_t^n being a permutation. Moreover, σ^n can be reconstructed from X^n and t_n .

Now consider a sequence of permutation processes $\{\sigma^n\}$. We are specifically interested in the case where $t_n \rightarrow \infty$. The limit of $\{\sigma^n\}$ is defined to be the weak limit of its associated trajectory processes as Borel probability measures on \mathbf{C} . As \mathbf{C} is a Polish space, the limit is also a Borel probability measure on \mathbf{C} . In other words, $\{\sigma^n\}$ converges if there is a stochastic process $X = (X(t), 0 \leq t \leq 1)$ with continuous sample paths such that for every uniformly continuous and bounded $F : \mathbf{C} \rightarrow \mathbb{R}$,

$$\mathbb{E}[F(X^n)] = \frac{1}{n} \sum_{i=1}^n F(T_i^n) \xrightarrow{n \rightarrow \infty} \mathbb{E}[F(X)]. \quad (1.5)$$

Our first result characterizes the limits of permutation processes.

Definition 1.1. A *permuton process* is a $[-1, 1]$ -valued stochastic process $X = (X(t), 0 \leq t \leq 1)$ with continuous sample paths and such that $X(t) \sim \text{Uniform}[-1, 1]$ for every t .

Theorem 1.1. For each n , let $\sigma^n = (\sigma_t^n; t \in [t_n])$ be a permutation process of \mathcal{S}_n with $t_n \rightarrow \infty$ as $n \rightarrow \infty$. Suppose $\{\sigma^n\}$ converges to a limit X in the sense of (1.5). Then X is a permuton process. Conversely, given any permuton process X there is a sequence of permutation processes that converges to X .

Theorem 1.1 extends the limit theory of single permutations to permutation processes. Indeed, if a sequence of permutation processes $\{\sigma^n\}$ has a limit X then for every $s \in [0, 1]$ the limit of $\sigma_{[s t_n]}^n$ is the permuton with the distribution of $(X(0), X(s))$. Moreover, for any set of times s_1, \dots, s_k , the empirical measure of the k -tuples $((i, \sigma_{[s_j t_n]}^n(i)); 1 \leq j \leq k)$ as i ranges over $[n]$ converges weakly, as measures rescaled onto $[-1, 1]^k$, to the distribution of $(X(s_1), \dots, X(s_k))$.

1.2 Random sorting networks

Recall from the previous section that a sorting network of \mathcal{S}_n is a shortest path from \mathbf{id}_n to \mathbf{rev}_n in the Cayley graph generated by adjacent transpositions. The number of sorting networks of \mathcal{S}_n was enumerated by Stanley [17], and later a combinatorial bijection with staircase shaped Young tableaux was provided by Edelman and Greene [6].

The number of permutations in a sorting network of \mathcal{S}_n is always $N := \binom{n}{2}$. A *random sorting network* of \mathcal{S}_n is a sorting network of \mathcal{S}_n chosen uniformly at random. We denote this random permutation process as $\mathbf{RSN}^n = (\mathbf{RSN}_k^n; 0 \leq k \leq N)$ (thus, \mathbf{RSN}_k^n is the k -th permutation in \mathbf{RSN}^n).

The asymptotic behaviour of \mathbf{RSN}^n was first studied by Angel et. al. [3]. It is shown that, as $n \rightarrow \infty$, the spacetime process of swaps of \mathbf{RSN}^n converges to the product of semicircle law and Lebesgue measure. It is also shown that, in the limit, the particle trajectories are Hölder-1/2 continuous, and the support of the permutation matrix lies within a certain octagon. Additional results about the asymptotic behaviour of \mathbf{RSN}^n have since been proved; see [1, 2] and the references therein. However, the main conjecture of [3], the Archimedean path conjecture, remains open. To state the conjecture and our results we first introduce the Archimedean measure.

The *Archimedean measure* is the unique probability measure on the plane with the property that all of its projections onto lines through the origin have the $\text{Uniform}[-1, 1]$ distribution. Its density, supported on the unit disk, is given by $(2\pi\sqrt{1-x^2-y^2})^{-1} dx dy$. It is also the projection of the normalized surface area measure of the Euclidean 2-sphere onto the unit disk. Let $(\mathbf{A}_x, \mathbf{A}_y)$ denote a random variable whose distribution is the Archimedean measure. The *Archimedean process* $\mathcal{A} = (\mathcal{A}(t); 0 \leq t \leq 1)$ is the permuton process defined by

$$\mathcal{A}(t) = \cos(\pi t) \mathbf{A}_x + \sin(\pi t) \mathbf{A}_y. \quad (1.6)$$

The Archimedean path conjecture [3, Conjecture 2] states that for every t the random permutation $\mathbf{RSN}_{[tN]}^n$ converges to the deterministic permuton $(\mathcal{A}(0), \mathcal{A}(t))$. The *Archimedean path* is the permuton valued path $\mathbf{A} = (\mathbf{A}(t); 0 \leq t \leq 1)$ such that

$$\mathbf{A}(t) \sim (\mathcal{A}(0), \mathcal{A}(t)) \text{ for every } 0 \leq t \leq 1. \quad (1.7)$$

Thus, the Archimedean path conjecture is that the empirical measures of permutations in \mathbf{RSN}^n converges to the Archimedean path. Figure 1 shows the support of the Archimedean path and the support of the empirical measures from a sample of \mathbf{RSN}^{500} at the same times.

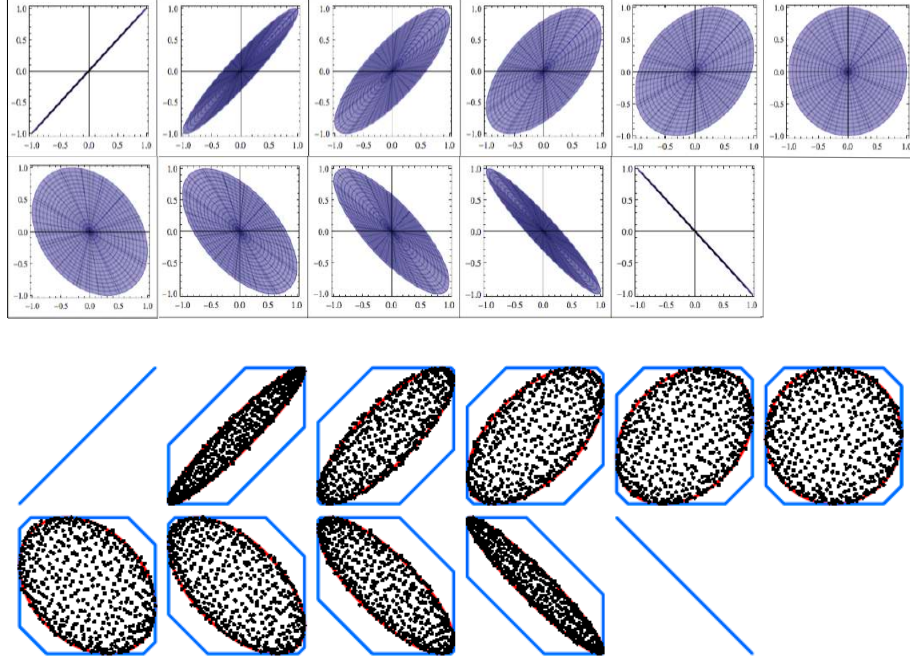


Figure 1: Support of the Archimedean path (top) and \mathbf{RSN}^{500} (bottom) at time intervals of $1/10$. Bottom figure is taken from [3, Figure 5].

Observe that the Archimedean process is a random sine curve. The trajectories of individual particles in \mathbf{RSN}^n also appear to be close to sine curves in simulations. Figure 2 shows the trajectories of particles in a sample from \mathbf{RSN}^{2000} . In fact, the sine curve conjecture [3, Conjecture 1] asserts that the trajectories of particles are close to random sine curves with high probability.

Based on the Archimedean path conjecture and the sine curve conjecture it is reasonable to conjecture the following.

Conjecture 1.1. \mathbf{RSN}^n converges in probability as a permutation process to the Archimedean process (1.6).

We emphasize that this conjecture states that the random trajectory process of \mathbf{RSN}^n concentrates around a deterministic limit and the limit is the Archimedean process. A similar phenomenon holds for the interchange process, which is explained in Section 2.1.

In addressing Conjecture 1.1 we provide a variational characterization of the Archimedean process. The (Dirichlet) energy of a stochastic process $X = (X(t); 0 \leq t \leq 1)$ is

$$\mathcal{E}[X] = \sup_{\Pi} \sum_{i=1}^k \frac{\mathbb{E}[|X(t_i) - X(t_{i-1})|^2]}{t_i - t_{i-1}},$$

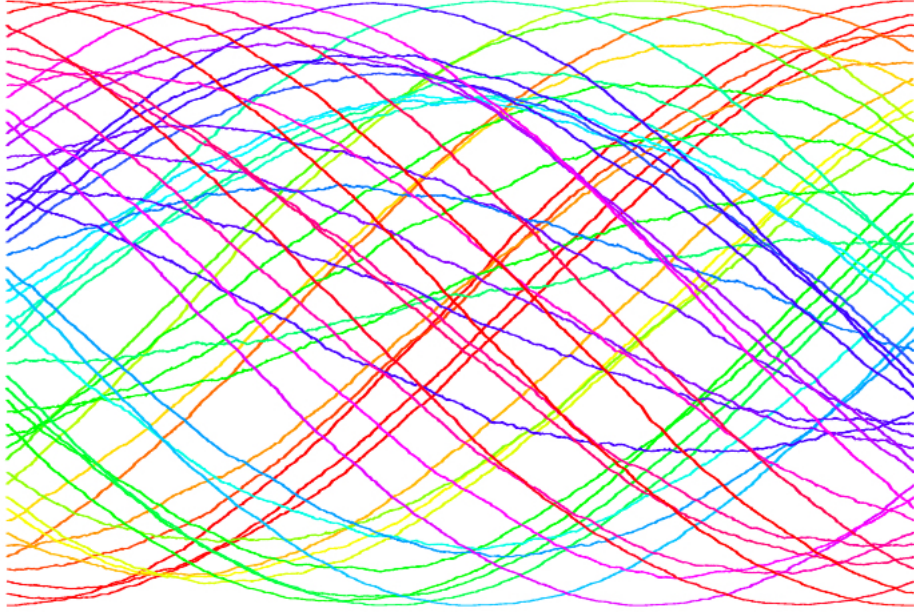


Figure 2: Some scaled particle trajectories from \mathbf{RSN}^{2000} . Taken from [3, Figure 1].

where the supremum is over all finite partitions $\Pi = \{0 = t_0 < t_1 < \dots < t_k = 1\}$ of $[0, 1]$. If X has continuously differentiable sample paths then $\mathcal{E}[X] = \int_0^1 \mathbb{E}[X'(t)^2] dt$. Thus, for example, a simple calculation shows that $\mathcal{E}[\mathcal{A}] = \pi^2/3$.

Theorem 1.2. *Among all permuton processes X with the property that $X(1) = -X(0)$, the Archimedean process \mathcal{A} uniquely minimizes the energy.*

The theorem allows for a characterization of stationary random permutation processes which converge to the Archimedean process in terms of the second moment of their speed. Random sorting networks are invariant under ϵ -shifts which take a trajectory $X(t)$ to $X(\epsilon + t)$. More precisely, the ϵ -shift makes the trajectory periodic with reversing boundary conditions, so the ϵ -shift is defined as

$$\left((-1)^{\lfloor \epsilon + t \rfloor} X(\epsilon + t \bmod 1), 0 \leq t \leq 1 \right).$$

We call a random trajectory process ϵ -stationary if its distribution is invariant under ϵ -shift of all individual particle trajectories.

Corollary 1.3. *Let X^n be a tight sequence of random trajectory processes that are ϵ_n -stationary with $\epsilon_n \rightarrow 0$. Assume that*

$$\limsup_{t \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{\mathbb{E}[(X^n(t) - X^n(0))^2]}{t^2} \leq \frac{\pi^2}{3}. \quad (1.8)$$

Then the sequence converges in probability to the deterministic limit given by the Archimedean process.

The trajectory process of \mathbf{RSN}^n is N^{-1} -stationary because of stationarity of the swaps of \mathbf{RSN}^n , [3, Theorem 1], which asserts that the random sequence of adjacent transpositions $(\tau_{s_1}, \dots, \tau_{s_N})$ that defines \mathbf{RSN}^n is stationary. Tightness of the random trajectory process of \mathbf{RSN}^n follows from [3,

Theorem 3] stating that for any $\delta > 0$, with probability tending to 1, all individual trajectories T^n of the particles in \mathbf{RSN}^n satisfy

$$|T^n(t) - T^n(s)| \leq \sqrt{8}|s - t|^{1/2} + \delta \text{ for every } s, t.$$

Corollary 1.3 is proved in Section 5.

Random sorting networks can also be studied in the setting of large deviation theory of the interchange process on paths. Consider the discrete time interchange process on the n -path, which is the path graph with n vertices. A random sorting network is the interchange process on the n -path conditioned to be at \mathbf{rev}_n in the shortest possible time N . Instead, one can consider *relaxed random sorting network*, which is the interchange process conditioned to be close to \mathbf{rev}_n in time $n^{2+\alpha}$ for some $\alpha \in (0, 1)$. One can study the relaxed network using large deviation theory in the following sense.

Suppose we fix a permuton process X that satisfies $X(0) = -X(1)$. One can ask what is the probability that the trajectory of a relaxed random sorting network on the n -path is close to X . This is the problem addressed in [11]. It is shown that this probability satisfies a large deviation principle whose rate function is the energy of X . Then Theorem 1.2 implies relaxed random sorting networks have to be close to the Archimedean process with high probability. This proves the Archimedean path conjecture for the relaxed networks.

Finally, we also provide a variational characterization of the Archimedean path in terms of energy in the 2-Wasserstein metric on the space of permutons. The 2-Wasserstein distance (henceforth, Wasserstein distance) between two Borel probability measures μ, ν on a metric space K is defined by

$$W(\mu, \nu)^2 = \inf_{\text{couplings } (V, W) \text{ s.t. } V \sim \mu, W \sim \nu} \mathbb{E} [d(V, W)^2]. \quad (1.9)$$

In order to study permutons in the Wasserstein metric we consider $K = [-1, 1]^2$ in the Euclidian metric. Let \mathbf{id} denote the identity permuton (X, X) and \mathbf{rev} denote the reverse permuton $(X, -X)$, where $X \sim \text{Uniform}[-1, 1]$.

Theorem 1.4. *Let $\mu = (\mu(t); 0 \leq t \leq 1)$ be a permuton valued path from $\mu(0) = \mathbf{id}$ to $\mu(1) = \mathbf{rev}$. Then the energy of μ in the Wasserstein metric satisfies $\mathcal{E}[\mu] \geq \mathcal{E}[\mathbf{A}] = \pi^2/6$, where \mathbf{A} is the Archimedean path (1.7). If there is equality then $\mu(t) = \mathbf{A}(t)$ for every t .*

A tool used in proving Theorem 1.4 may be of independent interest. We show that for a permuton valued path μ there exists a $[-1, 1]^2$ -valued stochastic process X such that the fixed time distributions of X is given by μ and the energy of X in the L^2 -metric equals the energy of μ in the Wasserstein metric. One may think of X as being an optimal coupling of the measures along μ . We prove such a ‘realization theorem’ for measure valued paths in a fairly general setting as stated in Theorem 4.1.

One motivation for Theorem 1.4 is that the Wasserstein distance is a natural distance on permutons. It is also related to sorting networks in the following way. A two-sided random sorting network is a shortest sequence of permutations from \mathbf{id}_n to \mathbf{rev}_n so that in each step the permutation is multiplied by an adjacent transposition either on left or on the right. This means that in each step, two adjacent columns or two adjacent rows of the permutation matrix are exchanged. Thus the 1s in the permutation matrix can be thought of as particles moving horizontally or vertically.

After scaling, we may consider the $[-1, 1]^2$ -valued trajectories for the n particles in a uniformly chosen two-sided sorting network. It can be shown that Conjecture 1.1 would imply that the trajectory process of two-sided random sorting networks converges to the optimal coupling of the Archimedean path **A**. See Figure 3 for the particle trajectories.

1.3 Permuton geometry

Let \mathcal{P} denote the space of all permutons. As in Theorem 1.4 it is natural to study \mathcal{P} in the Wasserstein metric. Motivated by Euclidean geometry, one may ask whether the sum of distances squared, $W(\mathbf{id}, \mathbf{P})^2 + W(\mathbf{P}, \mathbf{rev})^2$, is minimized by the Archimedean measure over all $\mathbf{P} \in \mathcal{P}$. The answer is false. Theorem 7.1 shows that this is minimized by a permuton \mathbf{P} if and only if \mathbf{P} remains a permuton after rotation by $\pi/4$ (!).

One can also ask whether there is a unique minimal energy path from \mathbf{id} to a given permuton \mathbf{P} . A motivation for this question is to understand minimal length paths from the identity to arbitrary permutations in the Cayley graph of \mathcal{S}_n generated by adjacent transpositions. These are called reduced decompositions. Counting reduced compositions is a deep and difficult combinatorial problem. We may get insights by studying related questions in the space of permutons. For instance, are there analogues of the Archimedean path conjecture for reduced decompositions of permutations approximating a target permuton \mathbf{P} ? Can large deviation theory provide an asymptotic count for the number of relaxed reduced decompositions of \mathbf{P} , à la sorting networks?

Theorem 7.4 proves that there is a minimal energy path from \mathbf{id} to a given permuton \mathbf{P} under some regularity condition on \mathbf{P} . This prompts us to pose the following problems.

Open problem 1.1 (Uniqueness of minimal energy paths). Under what conditions does there exist a unique minimal energy path in \mathcal{P} from \mathbf{id} to a given permuton \mathbf{P} ? What about for the Lebesgue permuton?

Open problem 1.2 (Diameter of permuton space). Suppose \mathbf{P} is a permuton such that there is a path of finite energy from \mathbf{id} to \mathbf{P} . Does the minimal energy path(s) to \mathbf{P} have energy at least that of the Archimedean path?

1.4 Outline of paper

We prove Theorem 1.1 in Section 2. In Section 3 we define path energy in metric spaces and discuss some of its basic properties. In Section 4 we prove Theorem 4.1 about realizing measure valued paths as stochastic processes. In Section 5 we prove Theorem 1.2 and Corollary 1.3. In Section 6 we prove Theorem 1.4. It is the most technical argument in the paper. Finally, in Section 7 we prove Theorem 7.1 and Theorem 7.4.

2 Limits of permutation processes

This section proves Theorem 1.1. We begin with a lemma about approximating continuous processes by their piecewise linear parts whose proof is in the Appendix.

Lemma 2.1. *Let $Y = (Y(t); 0 \leq t \leq 1)$ be a continuous $[-1, 1]$ -valued process. Consider its modulus of continuity $m^\delta(Y) = \sup_{s,t: |s-t| \leq \delta} |Y(s) - Y(t)|$. Then $\mathbb{E}[m^\delta(Y)] \rightarrow 0$ as $\delta \rightarrow 0$. Moreover, if Y and \hat{Y} are continuous processes then $|m^\delta(Y) - m^\delta(\hat{Y})| \leq 2\|Y - \hat{Y}\|_\infty$. Finally, let $\text{Lin}(n, Y)$ be the process obtained from Y such that it agrees with Y at times $t = i/n$ for $0 \leq i \leq n$ and is linear in between. Then $\mathbb{E}[\|\text{Lin}(n, Y) - Y\|_\infty] \rightarrow 0$ as $n \rightarrow \infty$.*

Limit of random trajectories are permuton processes. Suppose that the trajectory processes X^n of a sequence of permutation processes $(\sigma_t^n; t \in [t_n])$ converges to a continuous process X . We must show that $X(t) \sim \text{Uniform}[-1, 1]$ for every $t \in [0, 1]$. By Skorokhod's representation Theorem [9, Theorem 3.30] we may assume that the X^n and X are realized on a common probability space and $\|X^n - X\|_\infty \rightarrow 0$ almost surely. For a fixed t and every n consider the time $s_n \in \{i/t_n; 0 \leq i \leq t_n\}$ that is closest to t . Then $X^n(s_n)$ approximates $X(t)$ in distribution. Indeed,

$$|X(t) - X^n(s_n)| \leq |X(t) - X^n(t)| + |X^n(t) - X^n(s_n)|.$$

Now, $|X(t) - X^n(t)| \leq \|X - X^n\|_\infty$. Also, since $|t - s_n| \leq 1/t_n$, Lemma 2.1 implies

$$|X^n(t) - X^n(s_n)| \leq m^{1/t_n}(X^n) \leq m^{1/t_n}(X) + 2\|X - X^n\|_\infty.$$

Thus, $|X(t) - X^n(s_n)| \leq 3\|X^n - X\|_\infty + m^{1/t_n}(X)$. The term $m^{1/t_n}(X) \rightarrow 0$ almost surely in the sample outcomes of X due to almost sure continuity of X . Since $\|X^n - X\|_\infty \rightarrow 0$ almost surely, it follows that $|X(t) - X^n(s_n)| \rightarrow 0$ almost surely. The distribution of $X^n(s_n)$ is uniform on the set $\{\frac{2i}{n} - 1; i \in [n]\}$ as remarked earlier. Therefore, $X^n(s_n)$ converges weakly to $\text{Uniform}[-1, 1]$ and it follows that $X(t) \sim \text{Uniform}[-1, 1]$.

Permuton processes are limits of random trajectories Let $X = (X(t); 0 \leq t \leq 1)$ be the permuton process that is to be approximated by permutation processes. We will construct a sequence of random permutation processes and show that it converges almost surely to X . For $n \geq 1$ set $\Pi_n = \{i/n; 0 \leq i \leq n\}$. The following defines a random permutation process $(\sigma_t^n; t \in [n])$ of \mathcal{S}_n .

Let X_1, X_2, \dots be i.i.d. copies of X . For any fixed t the values $X_1(t), X_2(t), \dots$ are distinct almost surely. Therefore, we may assume that for every t in the countable set $\cup_n \Pi_n$ the values $X_1(t), X_2(t), \dots$ are all distinct. For each n and $t \in [n]$, the permutation σ_t^n is the ordering of the numbers $X_1(t/n), \dots, X_n(t/n)$ relative to the ordering of $X_1(0), \dots, X_n(0)$. In other words, consider the order statistics of the $X_k(0)$ s: $X_{(1)}(0) < X_{(2)}(0) < \dots < X_{(n)}(0)$. Let $\pi(i)$ be the index such that $X_{(i)}(0) = X_{\pi(i)}(0)$. Then for $i \in [n]$,

$$\sigma_t^n(i) = \text{rank of } X_{\pi(i)}(t/n) \text{ among } X_1(t/n), \dots, X_n(t/n).$$

Observe that π is a uniform random permutation that is measurable with respect to $(\text{w.r.t.}) X_1(0), \dots, X_n(0)$. Also, we naturally have $\sigma_0^n = \text{id}_n$ if we extend the definition above to $t = 0$.

Define the process Y_i by $Y_i(t) = (X_i(t) + 1)/2$. Then $Y_i(t) \sim \text{Uniform}[0, 1]$ for every t . For a fixed valued of $t \in [0, 1]$, set $\Delta_{i,j} = \mathbf{1}_{\{Y_j(t) \leq Y_i(t)\}} - Y_i(t)$. As the rank of a number y_i among y_1, \dots, y_n may

be expresses as $\sum_j \mathbf{1}_{\{y_j \leq y_i\}}$, we have that for any $t \in \Pi_n$,

$$\frac{2\sigma_{nt}^n(\pi^{-1}(i))}{n} - 1 - X_i(t) = \frac{2}{n} \sum_{j=1}^n \Delta_{i,j}. \quad (2.1)$$

Observe that $|\Delta_{i,j}| \leq 1$. Also, for $j \neq i$, $\mathbb{E}[\Delta_{i,j} \mid Y_i(t)] = 0$. This is where we use the fact that $X(t) \sim \text{Uniform}[-1, 1]$ for every t . Moreover, for all j such that $j \neq i$, the $\Delta_{i,j}$ s are mutually independent conditional on $Y_i(t)$. These observations allow us to use Bernstein's concentration inequality to bound $|\sum_{j:j \neq i} \Delta_{i,j}|$. We get that for $\epsilon \geq 0$ and every i ,

$$\mathbb{P} \left[\frac{1}{n-1} \left| \sum_{j:j \neq i} \Delta_{i,j} \right| > \epsilon \mid Y_i(t) \right] \leq 2e^{-\frac{\epsilon^2(n-1)}{4}}. \quad (2.2)$$

We use the above to bound the left hand side of (2.1) in probability. The right hand side (r.h.s.) of (2.1) is bounded in absolute value by $\frac{2}{n} + \frac{2|\sum_{j \neq i} \Delta_{i,j}|}{n-1}$. Therefore, taking an union bound over all $t \in \Pi_n$, setting $\epsilon = n^{-1/4}$ in (2.2) and then taking expectation over $Y_i(t)$ we infer that for all large n

$$\mathbb{P} \left[\sup_{t \in \Pi_n} \left| \frac{2\sigma_{nt}^n(\pi^{-1}(i))}{n} - 1 - X_i(t) \right| > 2n^{-1/4} + 2n^{-1} \right] \leq 2ne^{-\frac{n^{1/2}}{8}}.$$

Taking an union bound of the above over all particles i and reindexing $i \rightarrow \pi(i)$ we conclude

$$\mathbb{P} \left[\sup_{1 \leq i \leq n, t \in \Pi_n} \left| \frac{2\sigma_{nt}^n(i)}{n} - 1 - X_{\pi(i)}(t) \right| > 2n^{-1/4} + 2n^{-1} \right] \leq 2n^2 e^{-\frac{n^{1/2}}{8}}. \quad (2.3)$$

Recall that the trajectory of particle i is $T_i^n(t) = (2/n)\sigma_{nt}^n(i) - 1$ for $t \in \Pi_n$, and T_i^n is linearly interpolated in between the discrete times in Π_n . Let

$$A_n = \sup_{1 \leq i \leq n, t \in \Pi_n} |T_i^n(t) - X_{\pi(i)}(t)|.$$

Then $\mathbb{P}[A_n \geq 4n^{-1/4}]$ is summable by (2.3).

Let $\text{Lin}(n, X_i)$ be the piecewise linear function that agrees with X_i at times $t \in \Pi_n$. Observe that $\|T_i^n - \text{Lin}(n, X_{\pi(i)})\|_\infty = \sup_{t \in \Pi_n} |T_i^n(t) - X_{\pi(i)}(t)|$ because both functions are piecewise linear between the times $t \in \Pi_n$. Let

$$B_n = \frac{1}{n} \sum_i \|X_i - \text{Lin}(n, X_i)\|_\infty.$$

From another application of Bernstein's inequality, using that $\|X_i - \text{Lin}(n, X_i)\|_\infty \leq 2$ for all i , we deduce that

$$\mathbb{P} \left[B_n > \mathbb{E}[\|X - \text{Lin}(n, X)\|_\infty] + n^{-1/4} \right] \leq e^{-\frac{n^{1/2}}{16}}.$$

The r.h.s. of the above is summable over n and Lemma 2.1 also implies that $\mathbb{E}[\|X - \text{Lin}(n, X)\|_\infty] \rightarrow 0$ as $n \rightarrow \infty$. Furthermore, $\mathbb{P}[A_n \geq 4n^{-1/4}]$ is summable over n . Therefore, the Borel-Cantelli lemma implies that there is a subset Ω of outcomes of the X_i 's having probability one such that $A_n, B_n \rightarrow 0$ for every $\omega \in \Omega$.

Let $I \sim \text{Uniform}([n])$ and let \mathbb{E}_I denote expectation w.r.t. I , that is, the outcomes of the X_k s are kept fixed. Then,

$$\begin{aligned} \mathbb{E}_I[\|T_I^n - X_{\pi(I)}\|_\infty] &\leq \mathbb{E}_I[\|T_I^n - \text{Lin}(n, X_{\pi(I)})\|_\infty] + \mathbb{E}_I[\|\text{Lin}(n, X_{\pi(I)}) - X_{\pi(I)}\|_\infty] \\ &= \frac{1}{n} \sum_i \left(\sup_{t \in \Pi_n} |T_i^n(t) - X_{\pi(i)}(t)| \right) + \frac{1}{n} \sum_i \|\text{Lin}(n, X_i) - X_i\|_\infty \\ &\leq A_n + B_n. \end{aligned}$$

Therefore, if $\omega \in \Omega$ then $\mathbb{E}_I[\|T_I^n - X_{\pi(I)}\|_\infty] \rightarrow 0$. Consequently, for every uniformly continuous and bounded $F : \mathbf{C} \rightarrow \mathbb{R}$ we have that $\mathbb{E}_I[F(T_I^n) - F(X_{\pi(I)})] \rightarrow 0$ if $\omega \in \Omega$.

The distribution of $X_{\pi(I)}$ over the random I is the empirical measure $(1/n) \sum_i \delta_{X_i}$ on \mathbf{C} . As \mathbf{C} is a Polish space, the strong law of large numbers for empirical measures on Polish spaces [5, Theorem 11.4.1] implies that there is a subset Ω' of outcomes of the X_k s having probability one such that $X_{\pi(I)}$ converges weakly to X if $\omega \in \Omega'$.

If $\omega \in \Omega \cap \Omega'$ then for every function F as above we have that $\mathbb{E}_I[F(T_I^n)] \rightarrow \mathbb{E}[F(X)]$. Since $\mathbb{P}[\Omega \cap \Omega'] = 1$, we conclude that almost surely in the outcomes of the X_k s the function T_I^n converges weakly to X (the weak convergence is w.r.t. the random I). Since the distribution of T_I^n is the trajectory process of σ^n , we deduce that the sequence $\{\sigma^n\}$ converges almost surely to X . Selecting any such good outcome of the X_k s gives a deterministic sequence of permutation processes converging to X .

2.1 Convergence of random permutation processes

The limit notion for permutation processes automatically defines the limit notion for random permutation processes. More precisely, if σ^n is a random permutation process of \mathcal{S}_n then its trajectory process is a random measure on \mathbf{C} . Thus, a sequence of random permutation processes converges if their trajectory processes converge weakly as random measures on \mathbf{C} . In this case the limit is a priori a random permutation process, that is, a measure on permutation processes. This may seem a little strange so we discuss two contrasting examples.

First, we consider a sequence of random permutation processes that converge to a deterministic permutation process. The (discrete time) interchange process, Int^n , on the n -path is a random permutation process of \mathcal{S}_n generated by first sampling i.i.d. uniform random adjacent transpositions τ^1, τ^2, \dots , and then setting $\text{Int}_t^n = \text{Int}_{t-1}^n \circ \tau^t$ with $\text{Int}_0^n = \text{id}_n$. The stationary distribution of this process is the uniform measure of \mathcal{S}_n and its relaxation time is of order n^3 . Thus, we may expect a limit of the process if run until time n^3 .

This is indeed the case. It is shown in [15] that Int^n run until time n^3 converges to a deterministic permutation process: stationary Brownian motion on $[-1, 1]$. The latter is the law of standard Brownian motion on \mathbb{R} started from a uniform random point of $[-1, 1]$ and then reflected off of the horizontal lines $y = \pm 1$. ([15] considers the continuous time interchange process but the conclusion also holds for the discrete time process.) Observe that the trajectory of each particle of Int^n is a simple random walk on the n -path. So by Donsker's Theorem after appropriate rescaling it converges to Brownian motion on $[-1, 1]$ started from 0. However, this alone does not imply that the trajectory process itself converges to stationary Brownian motion. The convergence to the deterministic limit occurs

because the trajectories become *asymptotically independent*, which means the following. Let X_ω^n be the trajectory process for a sample outcome ω of Int^n . Let $T_{I_1}^n$ and $T_{I_2}^n$ be two samples from X_ω^n , that is, the trajectory of two particles I_1 and I_2 chosen independently and uniformly at random from X_ω^n . Then, for every continuous and bounded function $F : \mathbb{C} \rightarrow \mathbb{R}$,

$$\mathbb{E}_{\omega, I_1, I_2} [F(T_{I_1}^n)F(T_{I_2}^n)] - \mathbb{E}_{\omega, I_1} [F(T_{I_1}^n)]^2 \xrightarrow{n \rightarrow \infty} 0.$$

Asymptotic independence is the property that ensures that random permutation processes have deterministic limits.

The second example illustrates how lack of asymptotic independence provides random limits. Consider n particles placed on the vertices of the n -cycle, that is, n particles arranged in cyclic order. At each time step, rotate the cycle one unit clockwise or counter clockwise (by $2\pi/n$ radians) independently and uniformly at random. This gives a random permutation process σ^n whereby each particle performs a simple random walk on the n -cycle. However, the particle trajectories are not independent because the distances between particles remain fixed. When run until time n^3 this process has the following limit. Periodic Brownian motion on $[-1, 1]$, denoted B^{per} , is Brownian motion started from a uniform random point of $[-1, 1]$ and run in a period by identifying the endpoints ± 1 . Let $U \sim \text{Uniform}[-1, 1]$ be independent of B^{per} . The limit of $\{\sigma^n\}$ is the random permuton process $\omega \rightarrow X_\omega$ such that for a sample outcome U_ω of U , we have

$$X_\omega \stackrel{\text{law}}{=} U_\omega + B^{\text{per}} \pmod{[-1, 1]}.$$

In other words, $[-1, 1]$ is first rotated by U_ω (by identifying ± 1) and then rotated independently according to a periodic Brownian motion on $[-1, 1]$. Two samples from X_ω provide two periodic Brownian motions that start from a common point ω -almost surely.

A final remark is that the limit of the interchange process provides a derivation of the hydrodynamic limit of the simple symmetric exclusion process; see [15, Section 4].

3 Metric on permutons and path energy

This section provides the basic concepts used in order to prove the remaining main results of the paper. We work at some level of generality.

Let (K, d) be a metric space. A path $\gamma = (\gamma(t); 0 \leq t \leq 1)$ is a continuous function from the interval $[0, 1]$ into K . A *finite partition* of the interval $[a, b]$ is a set of ordered points $\Pi = \{a = t_0 < t_1 < \dots < t_n = b\}$. Let $\text{Part}[a, b]$ denote the set of all finite partitions of the interval $[a, b]$. The *mesh size* of a partition Π is $\Delta(\Pi) = \max_{1 \leq i \leq n} \{t_i - t_{i-1}\}$.

The energy of a path γ with respect to a partition $\Pi \in \text{Part}[0, 1]$ is

$$\mathcal{E}[\gamma, \Pi] = \sum_{i=1}^n \frac{d^2(\gamma(t_i), \gamma(t_{i-1}))}{t_i - t_{i-1}}.$$

The energy of γ , denoted $\mathcal{E}[\gamma]$, is

$$\mathcal{E}[\gamma] = \sup_{\Pi \in \text{Part}[0, 1]} \left\{ \mathcal{E}[\gamma, \Pi] \right\}. \quad (3.1)$$

The energy of γ restricted to the interval $[a, b]$ is

$$\mathcal{E}[\gamma, [a, b]] = \sup_{\Pi \in \text{Part}[a, b]} \left\{ \mathcal{E}[\gamma, \Pi] \right\}.$$

Notice that if $a \leq b \leq c$ then we have

$$\mathcal{E}[\gamma, [a, c]] \geq \mathcal{E}[\gamma, [a, b]] + \mathcal{E}[\gamma, [b, c]]. \quad (3.2)$$

In particular, if $t_0 \leq t_1 \leq \dots \leq t_n$ then $\mathcal{E}[\gamma, [t_0, t_n]] \geq \sum_{i=1}^n \mathcal{E}[\gamma, [t_{i-1}, t_i]]$.

For partitions Π, Π' of $[0, 1]$ we write $\Pi \subset \Pi'$ (Π' is a refinement of Π) if Π' contains all points of Π . The energy of a path is non-decreasing under refinements, as Lemma 3.1 below shows. The proof is in the Appendix. We will use this lemma throughout our arguments.

Lemma 3.1. *Suppose $\Pi \subset \Pi'$ are two finite partitions of $[0, 1]$. Then $\mathcal{E}[\gamma, \Pi] \leq \mathcal{E}[\gamma, \Pi']$ for any path γ in K .*

Using Lemma 3.1 we observe that for a path γ there is a sequence of nested finite partitions $\Pi_0 = \{0, 1\} \subset \Pi_1 \subset \Pi_2 \dots$ such that $\mathcal{E}[\gamma, \Pi_n] \nearrow \mathcal{E}[\gamma]$. We may also assume that $\Delta(\Pi_n) \rightarrow 0$. Thus $\cup_n \Pi_n$ is a dense set of points in $[0, 1]$.

We consider paths in two types of metric spaces. First, given a probability space (Ω, Σ, P) we take $K = L^2(P, [-1, 1])$, the Hilbert space of all square integrable random variables $Z : \Omega \rightarrow [-1, 1]$. A path in K is then a stochastic process $X = (X(t); 0 \leq t \leq 1)$. Permuton processes fall within this setup. Second, we take K to be the space of permutons \mathcal{P} in the Wasserstein metric, which is the setting of Theorem 1.4. The Wasserstein metric induces the topology of weak convergence on \mathcal{P} (see Lemma 8.1), and \mathcal{P} is compact in the weak topology by Phokhorov's Theorem.

It turns out that finite energy paths in the space of Borel probability measures of a compact metric space K are related to finite energy K -valued stochastic processes in that we may realize the former as the latter in an energy preserving manner. Here the probability measures are considered in the Wasserstein metric and the processes in the L^2 metric. This is the content of Theorem 4.1 in the following section 4, which is used to prove Theorem 1.4.

4 Realizing measure valued paths as stochastic processes

Throughout this section (K, d) denotes a compact metric space and $\mathcal{M}(K)$ denotes the space of Borel probability measures on K in the Wasserstein metric. A path $\gamma = (\gamma(t); 0 \leq t \leq 1)$ in $\mathcal{M}(K)$ is *realized* by a K -valued stochastic process $X = (X(t); 0 \leq t \leq 1)$ if the following conditions holds.

1. $X(t) \sim \gamma(t)$ for every t .
2. X has continuous sample paths almost surely.

The energy of X is as given by (3.1) with respect to the L^2 metric:

$$d_{L^2}(X(t), X(s)) := \mathbb{E} [d(X(t), X(s))^2]^{1/2}.$$

Theorem 4.1. *Suppose that γ is a $\mathcal{M}(K)$ -valued path with finite energy with respect to the Wasserstein distance. There is a stochastic process X that realizes γ in an energy preserving manner: $\mathcal{E}[X] = \mathcal{E}[\gamma]$.*

We refer to the process X as the optimal coupling of γ .

To begin the proof we note the following fact. If $\nu, \nu' \in \mathcal{M}(K)$ are two Borel probability measures then there is a coupling (V, W) of ν with ν' such that $W(\nu, \nu') = \mathbb{E} [d(V, W)^2]^{1/2}$. See Lemma 8.2 in the Appendix for proof. Using this fact we inductively build up optimal couplings by using the following lemma.

Lemma 4.2. *Let $\gamma_0, \dots, \gamma_n \in \mathcal{M}(K)$. There exist jointly distributed K -valued random variables (X_0, \dots, X_n) such that $X_i \sim \gamma_i$ and $\mathbb{E} [d(X_{i-1}, X_i)^2] = W(\gamma_{i-1}, \gamma_i)^2$ for $1 \leq i \leq n$.*

Proof. We proceed by induction. The case for two measures is mentioned above (see Lemma 8.2). To carry out the induction step we will need the following measure theoretic fact. It is often known as the Disintegration Theorem (see [9, Theorem 5.10]).

Fact: Let $(X, Y) \in K^2$ be jointly distributed random variables. There is a measurable function $g : K \times [0, 1] \rightarrow K^2$ such that if $U \sim \text{Uniform}[0, 1]$ and U is independent of (X, Y) then $(X, g(X, U))$ has the same joint distribution as (X, Y) .

Suppose the statement of the lemma holds for $\gamma_0, \dots, \gamma_{n-1}$ with jointly distributed random variables (X_0, \dots, X_{n-1}) . Using Lemma 8.2, we find a coupling (X'_{n-1}, X'_n) of γ_{n-1} with γ_n such that $W(\gamma_{n-1}, \gamma_n)^2 = \mathbb{E} [d(X'_{n-1}, X'_n)^2]$. Let $U \sim \text{Uniform}[0, 1]$ be independent of all the random variables $X_0, \dots, X_{n-1}, X'_{n-1}$ and X'_n . Let g be as mentioned in the fact above for the pair (X'_{n-1}, X'_n) .

Since X_{n-1} has the same distribution as X'_{n-1} , and U is independent of all other random variables, the pair $(X_{n-1}, g(X_{n-1}, U))$ has the same distribution as (X'_{n-1}, X'_n) . Let $X_n = g(X_{n-1}, U)$. Thus, $W(\gamma_{n-1}, \gamma_n)^2 = \mathbb{E} [d(X'_{n-1}, X'_n)^2] = \mathbb{E} [d(X_{n-1}, X_n)^2]$. The random variables (X_0, \dots, X_n) provide the desired coupling. \square

Proof of the Theorem 4.1

Suppose γ is a $\mathcal{M}(K)$ valued path. We may choose a sequence of nested finite partitions $\Pi_0 \subset \Pi_1 \dots$ such that $\Delta(\Pi_n) \rightarrow 0$ and $\mathcal{E}[\gamma, \Pi_n] \nearrow \mathcal{E}[\gamma]$.

For each n , we apply Lemma 4.2 to find coupled random variables $(X_n(t); t \in \Pi_n)$ such that if $\Pi_n = \{0 = t_0 < \dots < t_k = 1\}$ then

$$\mathbb{E} [d(X_n(t_i), X_n(t_{i-1}))^2] = W(\gamma(t_i), \gamma(t_{i-1}))^2 \text{ for every } 1 \leq i \leq k.$$

Fix an $x_0 \in K$. Set $\Pi_\infty = \cup_n \Pi_n$ and extend X_n to Π_∞ by setting $X_n(t) \equiv x_0$ if $t \in \Pi_\infty \setminus \Pi_n$.

The process X_n takes values in K^{Π_∞} for every n . As K^{Π_∞} is compact in the product topology, by applying Prokhorov's Theorem we can find a subsequence $n_i \rightarrow \infty$ and a process $(X(t); t \in \Pi_\infty)$ such that $X_{n_i} \rightarrow X$ weakly. As the partitions Π_n are nested we may assume w.l.o.g. that $n_i = n$, that is, $X_n \rightarrow X$ weakly.

Consider the process $(X(t); t \in \Pi_\infty)$. We must extend X continuously from the dense subset Π_∞ to $[0, 1]$. First, we show that X has finite energy along Π_∞ . Set $\mathcal{E}[X, \Pi_\infty] := \lim_{n \rightarrow \infty} \mathcal{E}[X, \Pi_n]$.

Lemma 4.3. *The process $(X(t); t \in \Pi_\infty)$ satisfies $\mathcal{E}[X, \Pi_\infty] \leq \mathcal{E}[\gamma]$. Moreover, $\mathbb{E}[d(X(t), X(s))^2] \leq (t-s)\mathcal{E}[\gamma, [s, t]]$ for every $s < t$ in Π_∞ .*

Proof. We begin by showing $\mathbb{E}[d(X(t), X(s))^2] \leq (t-s)\mathcal{E}[\gamma, [s, t]]$ for $s < t$ in Π_∞ . Suppose that $s < t$ are both in Π_∞ . From weak convergence of the X_n and compactness of K we have that $\mathbb{E}[d(X(t), X(s))^2] = \lim_{n \rightarrow \infty} \mathbb{E}[d(X_n(t), X_n(s))^2]$. We now bound $\mathbb{E}[d(X_n(t), X_n(s))^2]$. As $s, t \in \Pi_\infty$, there is an N such that $s, t \in \Pi_n$ for $n \geq N$. Suppose that the points of Π_n between s and t are $s = t_{0,n} < t_{1,n} < \dots < t_{k_n,n} = t$. Using Lemma 3.1 we deduce that for $n \geq N$,

$$\frac{\mathbb{E}[d(X_n(t), X_n(s))^2]}{t-s} \leq \sum_{i=1}^{k_n} \frac{\mathbb{E}[d(X_n(t_{n,i}), X_n(t_{n,i-1}))^2]}{t_{n,i} - t_{n,i-1}} = \sum_{i=1}^{k_n} \frac{W(\gamma(t_{n,i}), \gamma(t_{n,i-1}))^2}{t_{n,i} - t_{n,i-1}} \leq \mathcal{E}[\gamma, [s, t]].$$

The last inequality follows due to the $t_{n,i}$ forming a partition of $[s, t]$. By letting $n \rightarrow \infty$ we conclude from the above estimate that $\mathbb{E}[d(X(t), X(s))^2] \leq (t-s)\mathcal{E}[\gamma, [s, t]]$.

For the partition $\Pi_n = \{0 = t_0 < \dots < t_n = 1\}$ we deduce from the inequality above that

$$\sum_{i=1}^n \frac{\mathbb{E}[d(X(t_i), X(t_{i-1}))^2]}{t_i - t_{i-1}} \leq \sum_{i=1}^n \mathcal{E}[\gamma, [t_{i-1}, t_i]]. \quad (4.1)$$

From the inequality (3.2) we now deduce that $\sum_{i=1}^n \mathcal{E}[\gamma, [t_{i-1}, t_i]] \leq \mathcal{E}[\gamma]$. Therefore, $\mathcal{E}[X, \Pi_\infty] \leq \mathcal{E}[\gamma]$ as required. \square

We now show that X has a continuous extension to a process defined for times $t \in [0, 1]$. Let (Ω, Σ, μ) denote the probability space where $(X(t), t \in \Pi_\infty)$ is jointly defined and let $X_\omega(t)$ denote the outcome of $X(t)$ for $\omega \in \Omega$. The inequality $\mathcal{E}[X, \Pi_\infty] \leq \mathcal{E}[\gamma]$ from Lemma 4.3 implies that for μ -almost every ω the energy of the discrete K -valued path $(X_\omega(t), t \in \Pi_\infty)$ is finite. In other words, for μ -almost every ω ,

$$\sup_n \sum_{t_i \in \Pi_n} \frac{d^2(X_\omega(t_i), X_\omega(t_{i-1}))}{t_i - t_{i-1}} < \infty.$$

In particular, for μ -almost every ω there exists a constant C_ω such that

$$d(X_\omega(t), X_\omega(s)) \leq C_\omega \sqrt{|t-s|} \text{ for } s, t \in \Pi_\infty.$$

Since Π_∞ is a dense subset of $[0, 1]$, Lemma 8.3 from the Appendix implies that $(X_\omega(t); t \in \Pi_\infty)$ has a continuous extension to times $t \in [0, 1]$ for μ -almost every ω . We denote this extension for μ -almost every ω by $X = (X(t), 0 \leq t \leq 1)$, which is then a K -valued stochastic process with continuous sample paths.

Now we show that X realizes γ . Certainly, $X(t) \sim \gamma(t)$ for $t \in \Pi_\infty$ because $X_n(t) \rightarrow X(t)$ weakly and $X_n(t) \sim \gamma(t)$ for all large n due to the partitions Π_n being nested. Suppose that $t \in [0, 1] \setminus \Pi_\infty$. Choose a sequence $t_n \in \Pi_n$ such that $t_n \rightarrow t$. By continuity of X and the bounded convergence theorem we conclude that $\mathbb{E}[d(X(t_n), X(t))] \rightarrow 0$. This implies that $X(t_n) \rightarrow X(t)$ weakly. The distribution of $X(t_n)$ is $\gamma(t_n)$ and $\gamma(t_n) \rightarrow \gamma(t)$ weakly because the path γ is continuous due to having finite energy. Therefore, $X(t) \sim \gamma(t)$ and X realizes γ .

Finally we show that $\mathcal{E}[X] = \mathcal{E}[\gamma]$. As X realizes γ , $\mathbb{E}[d(X(t), X(s))^2] \geq W(\gamma(t), \gamma(s))^2$. Hence,

$\mathcal{E}[X] \geq \mathcal{E}[\gamma]$. To get the reverse inequality first recall from Lemma 4.3 that

$$\mathbb{E} [d(X(t), X(s))^2] \leq (t-s) \mathcal{E}[\gamma, [s, t]] \text{ for every } s, t \in \Pi_\infty \text{ with } s < t.$$

Suppose $s < t$ are two arbitrary points in $[0, 1]$. Choose sequences $\{s_n\}$ and $\{t_n\}$ such that $s_n, t_n \in \Pi_n$, $s_n \leq t_n$, $s_n \searrow s$ and $t_n \nearrow t$. From continuity of X and the bounded convergence theorem we have that $\mathbb{E} [d(X(t), X(s))^2] = \lim_{n \rightarrow \infty} \mathbb{E} [d(X(t_n), X(s_n))^2]$. Since

$$\mathbb{E} [d(X(t_n), X(s_n))^2] \leq (t_n - s_n) \mathcal{E}[\gamma, [s_n, t_n]] \text{ and } \mathcal{E}[\gamma, [s_n, t_n]] \leq \mathcal{E}[\gamma, [s, t]],$$

we conclude that $\mathbb{E} [d(X(t), X(s))^2] \leq (t-s) \mathcal{E}[\gamma, [s, t]]$ for every $s \leq t$. For an arbitrary partition $\Pi = \{0 = t_0 < \dots < t_n = 1\}$ we use this inequality to deduce that $\mathcal{E}[X, \Pi] \leq \sum_{i=1}^n \mathcal{E}[\gamma, [t_{i-1}, t_i]]$. Inequality (3.2) implies that $\sum_{i=1}^n \mathcal{E}[\gamma, [t_{i-1}, t_i]] \leq \mathcal{E}[\gamma]$. As Π was arbitrary it follows that $\mathcal{E}[X] \leq \mathcal{E}[\gamma]$. This completes the proof.

5 Minimal energy permuton processes from identity to reverse

In this section we prove Theorem 1.2 and Corollary 1.3. The proof of Theorem 1.2 uses the following lemma about minimal energy paths on a Hilbert sphere. A proof of the lemma is provided in the Appendix.

Lemma 5.1. *Let γ be a path on the unit sphere of a Hilbert space between two antipodal points $\gamma(0)$ and $-\gamma(0) = \gamma(1)$. Then $\mathcal{E}[\gamma] \geq \pi^2$ with equality if and only if*

$$\gamma(t) = \cos(\pi t)\gamma(0) + \sin(\pi t)\gamma(1/2).$$

Proof of Theorem 1.2. Suppose $X = (X(t); 0 \leq t \leq 1)$ is a permuton process with $X(0) = -X(1)$. Since $\mathbb{E} [X(t)^2] = 1/3$, the process X is a path between two antipodal points on the sphere of radius $1/\sqrt{3}$ in the Hilbert space $L^2(\Omega, \Sigma, P)$, where (Ω, Σ, P) is the probability space over which the process X is defined. From Lemma 5.1 we see that $\mathcal{E}[X] \geq \pi^2/3$ with equality if and only if $X(t) = \cos(\pi t)X(0) + \sin(\pi t)X(1/2)$. In case of equality, since $X(t) \sim \text{Uniform}[-1, 1]$ for every t , this equation for X implies that the pair $(X(0), X(1/2))$ has uniform projections and must be distributed according to the Archimedean measure. Consequently, X is the Archimedean process. \square

Proof of Corollary 1.3. Theorem 1.2 implies that all limit points of X^n are supported on permuton processes with energy at least $\pi^2/3$. Due to the uniqueness of energy minimizers, X^n will converge to the Archimedean process if the expected energy of any limit point X of X^n is at most $\pi^2/3$. By Fatou's Lemma and assumption (1.8) we have

$$\mathbb{E} [(X(t) - X(0))^2] \leq \frac{(\pi t)^2}{3}.$$

The ϵ_n -stationarity of X^n implies ϵ -stationarity of X for all $\epsilon > 0$, so

$$\mathbb{E}[(X(t) - X(s))^2] \leq \frac{(\pi(s-t))^2}{3},$$

and therefore for every partition $\Pi = \{t_0 = 0, \dots, t_k = 1\}$ we have

$$\sum_{i=1}^k \frac{\mathbb{E}[(X(t_i) - X(t_{i-1}))^2]}{t_i - t_{i-1}} \leq \frac{\pi^2}{3}. \quad \square$$

Note that conversely, the Archimedean process limit and the bounded convergence theorem would imply that for every t , $\mathbb{E}[(X^n(t) - X^n(0))^2] \rightarrow \frac{2}{3}(1 - \cos(\pi t))$.

6 Minimal energy permuton paths from identity to reverse

In this section we prove Theorem 1.4. As there are a few steps we first outline the strategy. Using Theorem 4.1 we transfer the study of paths in \mathcal{P} to $[-1, 1]^2$ -valued stochastic processes. Then we solve the corresponding energy minimization problem for stochastic processes. We verify that there is an unique energy minimizer and its fixed time marginals agree with the Archimedean path. In the following we begin by describing an optimal coupling of the Archimedean path, which will be helpful with the aforementioned verification step.

Lemma 6.1. *For $0 \leq s < t \leq 1$, $W(\mathbf{A}(s), \mathbf{A}(t)) = \sqrt{\frac{8}{3}} \sin(\frac{\pi}{4}(t-s))$. As a result, $\mathcal{E}[\mathbf{A}] = \pi^2/6$. Furthermore, an optimal coupling of the Archimedean path is as follows. Let $(\mathbf{A}_x, \mathbf{A}_y)$ be distributed according to the Archimedean measure. The optimal coupling $\mathbf{P}(t) = (\mathbf{P}_x(t), \mathbf{P}_y(t))$ is*

$$\begin{aligned} \mathbf{P}_x(t) &= \cos\left(\frac{\pi}{2}t\right) \mathbf{A}_x + \sin\left(\frac{\pi}{2}t\right) \mathbf{A}_y \\ \mathbf{P}_y(t) &= \cos\left(\frac{\pi}{2}t\right) \mathbf{A}_x - \sin\left(\frac{\pi}{2}t\right) \mathbf{A}_y. \end{aligned} \quad (6.1)$$

Proof. The map $t \rightarrow \mathbf{P}(t)$ is a parametrization of the Archimedean path. One can verify that $\mathbf{P}(t)$ and $\mathbf{A}(t)$ have the same density by computing their densities from the density of the Archimedean measure. As $(\mathbf{P}(s), \mathbf{P}(t))$ is a coupling of $\mathbf{A}(s)$ with $\mathbf{A}(t)$, $W(\mathbf{A}(s), \mathbf{A}(t))^2 \leq \mathbb{E}[\|\mathbf{P}(t) - \mathbf{P}(s)\|^2]$. Since $\mathbb{E}[\mathbf{A}_x \mathbf{A}_y] = 0$ and $\mathbb{E}[\mathbf{A}_x^2] = \mathbb{E}[\mathbf{A}_y^2] = 1/3$,

$$\mathbb{E}[\|\mathbf{P}(t) - \mathbf{P}(s)\|^2] = \frac{2}{3} \left[\left(\cos\left(\frac{\pi}{2}t\right) - \cos\left(\frac{\pi}{2}s\right) \right)^2 + \left(\sin\left(\frac{\pi}{2}t\right) - \sin\left(\frac{\pi}{2}s\right) \right)^2 \right].$$

The sum of squares following the factor of $2/3$ is the squared Euclidian distance between the vectors $(\cos(\frac{\pi}{2}s), \sin(\frac{\pi}{2}s))$ and $(\cos(\frac{\pi}{2}t), \sin(\frac{\pi}{2}t))$. It equals $4\sin(\frac{\pi}{4}(t-s))^2$. This establishes that $W(\mathbf{A}(t), \mathbf{A}(s)) \leq \sqrt{8/3} \sin(\frac{\pi}{4}(t-s))$.

Now we prove the required lower bound for $W(\mathbf{A}(s), \mathbf{A}(t))$. The idea is as follows. Consider any coupling (V, W) of $\mathbf{A}(t)$ with $\mathbf{A}(s)$. We rotate both V and W by $\pi/4$ radians and consider the vector $A = (\frac{V_1 - V_2}{\sqrt{2}}, \frac{V_1 + V_2}{\sqrt{2}})$ and the corresponding rotated vector B for W .

The distribution of $\frac{V_1 - V_2}{\sqrt{2}}$ depends only on $\mathbf{A}(t)$ and not on the coupling V . In fact, it is distributed as $\text{Uniform}[-\sqrt{2} \sin(\frac{\pi}{2}t), \sqrt{2} \sin(\frac{\pi}{2}t)]$, which can be deduced by calculating the distribution of $\frac{\mathbf{P}_x(t) - \mathbf{P}_y(t)}{\sqrt{2}}$. Similarly, $\frac{V_1 + V_2}{\sqrt{2}} \sim \text{Uniform}[-\sqrt{2} \cos(\frac{\pi}{2}t), \sqrt{2} \cos(\frac{\pi}{2}t)]$. Therefore,

$$\begin{aligned}
\mathbb{E} [||V - W||^2] &= \mathbb{E} [||A - B||^2] \\
&= \mathbb{E} \left[\left| \frac{V_1 - V_2}{\sqrt{2}} - \frac{W_1 - W_2}{\sqrt{2}} \right|^2 \right] + \mathbb{E} \left[\left| \frac{V_1 + V_2}{\sqrt{2}} - \frac{W_1 + W_2}{\sqrt{2}} \right|^2 \right] \\
&\geq W(U[\sqrt{2} \sin(\frac{\pi}{2}t)], U[\sqrt{2} \sin(\frac{\pi}{2}s)])^2 + W(U[\sqrt{2} \cos(\frac{\pi}{2}t)], U[\sqrt{2} \cos(\frac{\pi}{2}s)])^2,
\end{aligned}$$

where $U[a]$ denotes the Uniform $[-a, a]$ distribution. The inequality above implies

$$W(\mathbf{A}(t), \mathbf{A}(s))^2 \geq W(U[\sqrt{2} \sin(\frac{\pi}{2}t)], U[\sqrt{2} \sin(\frac{\pi}{2}s)])^2 + W(U[\sqrt{2} \cos(\frac{\pi}{2}t)], U[\sqrt{2} \cos(\frac{\pi}{2}s)])^2. \quad (6.2)$$

Now, $W(U[a], U[b])^2 = \frac{(b-a)^2}{3}$. The optimal coupling is (aU, bU) with $U \sim \text{Uniform}[-1, 1]$. Substituting this into the r.h.s. of the inequality from (6.2) and simplifying we deduce that $W(\mathbf{A}(t), \mathbf{A}(s)) \geq \sqrt{\frac{8}{3}} \sin(\frac{\pi}{4}(t-s))$.

The idea of rotating measures to get a lower bound on the Wasserstein distance will be used in the proof of Theorem 1.4 as well. \square

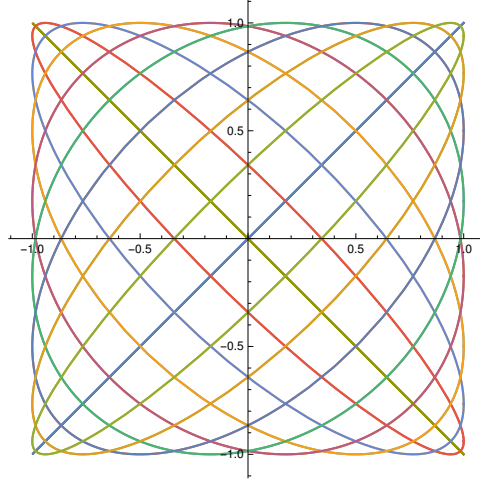


Figure 3: Selected trajectories of points, or particles, along the unit circle under the optimal coupling of the Archimedean path (6.1). The point $(\cos(\theta), \sin(\theta))$ traces out a quarter arc of an ellipse starting at $(\cos(\theta), \cos(\theta))$ and ending at $(\sin(\theta), -\sin(\theta))$. Trajectories of points with angles $\theta, \pi - \theta, \pi + \theta$ and $2\pi - \theta$ trace out an ellipse.

1-dimensional energy minimization In the proof of Theorem 1.4 we will reduce the 2-dimensional energy minimization problem involving $[-1, 1]^2$ -valued stochastic processes to a pair of 1-dimensional energy minimization problem. Here we solve that 1-dimensional problem.

For continuous $f : [0, 1] \rightarrow \mathbb{R}$, let $\mathcal{E}[f]$ and $\mathcal{E}[f, \Pi]$ denote energy w.r.t. the Euclidian metric on \mathbb{R} .

Lemma 6.2. *Let $X = (X(t); 0 \leq t \leq 1)$ be a continuous \mathbb{R} -valued stochastic process such that $\mathbb{E}[X(t)] = 0$ and $\mathbb{E}[X(t)^2] < \infty$ for every t . Set $\sigma(t) = \mathbb{E}[X(t)^2]^{1/2}$. Then $\mathcal{E}[X] \geq \mathcal{E}[\sigma]$. Here the*

energy of X is w.r.t. the L^2 -metric and the energy of σ is w.r.t. the Euclidian metric.

Moreover, suppose that $\mathcal{E}[X] = \mathcal{E}[\sigma]$, $\mathcal{E}[\sigma] < \infty$ and $\sigma(t) > 0$ for $t > 0$. Then the following holds almost surely,

$$X(t) = \frac{\sigma(t)}{\sigma(1)} X(1) \text{ for every } 0 \leq t \leq 1.$$

Proof. We have that $\mathbb{E}[|X(t) - X(s)|^2] = \sigma(t)^2 - 2\mathbb{E}[X(t)X(s)] + \sigma(s)^2$. The Cauchy-Schwarz inequality implies $\mathbb{E}[X(t)X(s)] \leq \sigma(t)\sigma(s)$, and hence, $\mathbb{E}[|X(t) - X(s)|^2] \geq (\sigma(t) - \sigma(s))^2$. From this inequality it is immediate that $\mathcal{E}[X] \geq \mathcal{E}[\sigma]$.

Now suppose that $\mathcal{E}[X] = \mathcal{E}[\sigma]$, $\mathcal{E}[\sigma]$ is finite and $\sigma(t) > 0$ for every $t > 0$. If we show that $X(t)/\sigma(t)$ is almost surely constant on the interval $[\epsilon, 1]$, for any $\epsilon > 0$, then the continuity of X implies that $X(t)/\sigma(t)$ is almost surely constant on $[0, 1]$. Therefore, we may assume that $\sigma(t) > 0$ for $t \in [0, 1]$.

Set $\delta = \inf_{t \in [0, 1]} \sigma(t)$. Then $\delta > 0$ since $\sigma(t)$ is continuous and positive on $[0, 1]$. Set $Y(t) = \frac{X(t)}{\sigma(t)}$. For $0 \leq s \leq t \leq 1$,

$$\mathbb{E}[|Y(t) - Y(s)|^2] = \frac{\mathbb{E}[|X(t) - X(s)|^2] - |\sigma(t) - \sigma(s)|^2}{\sigma(s)\sigma(t)} \leq \frac{\mathbb{E}[|X(t) - X(s)|^2] - |\sigma(t) - \sigma(s)|^2}{\delta^2}.$$

The estimate above implies that for any finite partition Π of $[0, 1]$,

$$\mathcal{E}[Y, \Pi] \leq \delta^{-2}(\mathcal{E}[X, \Pi] - \mathcal{E}[\sigma, \Pi]) \leq \delta^{-2}(\mathcal{E}[X] - \mathcal{E}[\sigma, \Pi]).$$

Choose a sequence of nested partitions $\Pi_0 \subset \Pi_1 \subset \dots$ such that $\mathcal{E}[\sigma, \Pi_n] \rightarrow \mathcal{E}[\sigma]$. We deduce from the above that $\mathcal{E}[Y, \Pi_n] \rightarrow 0$ due to $\mathcal{E}[X] = \mathcal{E}[\sigma]$. Since $\mathcal{E}[Y, \Pi_n]$ is monotone increasing we conclude that $\mathcal{E}[Y, \Pi_n] = 0$ for every n . Set $\Pi = \cup_n \Pi_n$. Then for every $s, t \in \Pi$,

$$\mathbb{E}[|Y(t) - Y(s)|^2] \leq |t - s| \cdot \left(\sup_n \mathcal{E}[Y, \Pi_n] \right) = 0.$$

Fix an arbitrary $p \in \Pi$. We deduce from the above that for every $q \in \Pi$, $\mathbb{P}[Y(q) = Y(p)] = 1$. Taking the countable intersection of these events over all $q \in \Pi$ we conclude that

$$\mathbb{P}[Y(q) = Y(p) \text{ for every } q \in \Pi] = 1.$$

The continuity of Y and the fact that Π is a dense subset of $[0, 1]$ imply that almost surely, $Y(t) \equiv Y(p)$ for every $t \in [0, 1]$. In other words, $X(t) = \frac{\sigma(t)}{\sigma(1)} X(1)$ for every t , almost surely, as required. \square

6.1 Proof of Theorem 1.4

Let $\mathbf{P} = (\mathbf{P}(t); 0 \leq t \leq 1)$ be a path in \mathcal{P} from **id** to **rev** such that $\mathcal{E}[\mathbf{P}] < \infty$ in the Wasserstein metric. From Theorem 4.1 we may realize \mathbf{P} as a $[-1, 1]^2$ -valued continuous stochastic process X such that $\mathcal{E}[\mathbf{P}] = \mathcal{E}[X]$.

Write $X(t) = (x(t), y(t))$. Then $x(t)$ and $y(t)$ are distributed as Uniform $[-1, 1]$ since $\mathbf{P}(t)$ is a

permuton. Also, $x(0) = y(0)$ and $x(1) = -y(1)$ due to $\mathbf{P}(0) = \mathbf{id}$ and $\mathbf{P}(1) = \mathbf{rev}$. Set

$$\begin{aligned} u(t) &= \frac{x(t) - y(t)}{\sqrt{2}}, \\ v(t) &= \frac{x(t) + y(t)}{\sqrt{2}}. \end{aligned}$$

Then $\mathbb{E}[u(t)] = \mathbb{E}[v(t)] = 0$ for every t . For the boundary conditions we have $u(0) = 0$ and $u(1) = \sqrt{2}x(1)$, while $v(0) = \sqrt{2}x(0)$ and $v(1) = 0$. Set

$$\sigma_u^2(t) = \mathbb{E}[u(t)^2] \quad \text{and} \quad \sigma_v^2(t) = \mathbb{E}[v(t)^2].$$

Since $u(t)^2 + v(t)^2 = x(t)^2 + y(t)^2$, we see that $\sigma_u^2(t) + \sigma_v^2(t) = \mathbb{E}[x(t)^2 + y(t)^2] = 2/3$ due to $x(t)$ and $y(t)$ being distributed as $\text{Uniform}[-1, 1]$.

The map $t \rightarrow (\sigma_u(t), \sigma_v(t))$ is a path on the circle of radius $\sqrt{2/3}$ that begins at $(0, \sqrt{2/3})$ and ends at $(\sqrt{2/3}, 0)$. It is well known that there is a unique path of minimal energy on such a circle from $(0, \sqrt{2/3})$ to $(\sqrt{2/3}, 0)$. This is the minor arc going from $(0, \sqrt{2/3})$ to $(\sqrt{2/3}, 0)$, and uniquely parametrized by $t \rightarrow \sqrt{2/3}(\sin(\frac{\pi}{2}t), \cos(\frac{\pi}{2}t))$. (We actually show this during the proof of Lemma 5.1 and the reader may also see [13, chapter 5].) The energy of this path is

$$\frac{\pi^2}{6} \int_0^1 \cos\left(\frac{\pi}{2}t\right)^2 + \sin\left(\frac{\pi}{2}t\right)^2 dt = \frac{\pi^2}{6}.$$

Consequently, $\mathcal{E}[(\sigma_u, \sigma_v), \Pi]^2 \geq \pi^2/6$. In case of equality we must have

$$\sigma_u(t) = \sqrt{2/3} \sin\left(\frac{\pi}{2}t\right) \quad \text{and} \quad \sigma_v(t) = \sqrt{2/3} \cos\left(\frac{\pi}{2}t\right). \quad (6.3)$$

Note that $\mathbb{E}[|X(t) - X(s)|^2] = |u(t) - u(s)|^2 + |v(t) - v(s)|^2$, where the distance for X is in the Euclidian metric of \mathbb{R}^2 . This implies that $\mathcal{E}[X] = \mathcal{E}[u] + \mathcal{E}[v]$. Lemma 6.2 then implies $\mathcal{E}[u] + \mathcal{E}[v] \geq \mathcal{E}[\sigma_u] + \mathcal{E}[\sigma_v]$. Therefore,

$$\mathcal{E}[\gamma] = \mathcal{E}[X] = \mathcal{E}[u] + \mathcal{E}[v] \geq \mathcal{E}[\sigma_u] + \mathcal{E}[\sigma_v].$$

However, $\mathcal{E}[\sigma_u] + \mathcal{E}[\sigma_v] = \mathcal{E}[(\sigma_u, \sigma_v)] \geq \pi^2/6$.

We have deduced that $\mathcal{E}[\gamma] \geq \pi^2/6$ for any path γ in \mathcal{P} . As $\mathcal{E}[\mathbf{A}] = \pi^2/6$, we deduce that $(\mathbf{A}(t); 0 \leq t \leq 1)$ has minimal energy among all paths from \mathbf{id} to \mathbf{rev} in \mathcal{P} . If $\mathcal{E}[\gamma] = \pi^2/6$ then the functions σ_u and σ_v must equal the functions from (6.3). In this case we may apply the case of equality from Lemma 6.2 to the processes $u(t)$ and $v(t)$ to conclude that $u(t) = \sqrt{2} \sin(\frac{\pi}{2}t) x(1)$ and $v(t) = \sqrt{2} \cos(\frac{\pi}{2}t) x(0)$ for every t , almost surely. In terms of X we get that almost surely, for all $0 \leq t \leq 1$,

$$\begin{aligned} x(t) &= \cos\left(\frac{\pi}{2}t\right) x(0) + \sin\left(\frac{\pi}{2}t\right) x(1) \\ y(t) &= \cos\left(\frac{\pi}{2}t\right) x(0) - \sin\left(\frac{\pi}{2}t\right) x(1). \end{aligned} \quad (6.4)$$

We claim that (6.4) implies $(x(0), x(1))$ is distributed as the Archimedean measure. If this holds

then we have $\gamma(t) = \mathbf{A}(t)$ because $X(t) \sim \gamma(t)$ and (6.4) implies that $X(t) \sim \mathbf{A}(t)$. We have seen the latter fact in (6.1) whereby the r.h.s. of (6.4) has the $\mathbf{A}(t)$ distribution if $(x(0), x(1))$ is distributed according to the Archimedean measure.

To see that $(x(0), x(1))$ is distributed as the Archimedean measure observe that $x(t)$ is the projection of $(x(0), x(1))$ onto the line through the origin with angle $\frac{\pi}{2}t$. Also, $y(t)$ is the projection of $(x(0), x(1))$ on the line through the origin with angle $-\frac{\pi}{2}t$. As $x(t)$ and $y(t)$ are distributed as Uniform $[-1, 1]$ for every t , it follows that the distribution of the projection of $(x(0), x(1))$ onto any line through the origin is Uniform $[-1, 1]$. This property determines the Archimedean measure. This completes the proof of Theorem 1.4.

7 Geometry of permutons

7.1 Permuton space is not spherical

It may appear that \mathcal{P} resembles a sphere in the Wasserstein metric, however, this is far from the case as the next theorem shows.

Theorem 7.1. *The function $\mathbf{P} \rightarrow W(\mathbf{id}, \mathbf{P})^2 + W(\mathbf{P}, \mathbf{rev})^2$ is minimized by a permuton $\mathbf{P} \sim (X, Y)$ if and only if the pair $(\frac{X-Y}{\sqrt{2}}, \frac{X+Y}{\sqrt{2}})$ is also a permuton. In particular, the Archimedean measure is not the unique minimizer.*

The proof of the following lemma, used in proving Theorem 7.1, is in the Appendix.

Lemma 7.2. *Let σ and τ be permutations in \mathcal{S}_n . Then,*

$$W(\mu_\sigma, \mu_\tau)^2 = \frac{4}{n^3} \inf_{\pi \in \mathcal{S}_n} \sum_i (i - \pi(i))^2 + (\sigma(i) - \tau(\pi(i)))^2.$$

Lemma 7.3. *The Wasserstein distance of the identity permuton \mathbf{id} from any permuton $\mathbf{P} = (X, Y)$ is expressed as follows. Let (X', Y') denote an independent copy of (X, Y) . Then,*

$$W(\mathbf{id}, \mathbf{P})^2 = \frac{4}{3} - 2\mathbb{E}[\max\{X + Y, X' + Y'\}].$$

Proof. We first derive the analogue of the above formula for permutations and then take limits to get the final result. For a permutation $\sigma \in \mathcal{S}_n$, consider

$$\sum_i (i - \pi(i))^2 + (i - \sigma\pi(i))^2 = 4 \sum_i i^2 - 2 \sum_i i \cdot (\pi(i) + \sigma\pi(i)).$$

Reindexing the latter sum by setting $i = \pi^{-1}(i)$ and replacing π by π^{-1} we get that

$$\inf_{\pi \in \mathcal{S}_n} \sum_i (i - \pi(i))^2 + (i - \sigma\pi(i))^2 = 4 \sum_i i^2 - 2 \sup_{\pi} \sum_i \pi(i)(i + \sigma(i)).$$

The sum $\sum_i \pi(i) \cdot (i + \sigma(i))$ is maximized by choosing $\pi(i)$ to be the rank of $i + \sigma(i)$ in the sequence $1 + \sigma(1), \dots, n + \sigma(n)$. Then π satisfies $\pi(1) + \sigma(\pi(1)) \leq \pi(2) + \sigma(\pi(2)) \leq \dots \leq \pi(n) + \sigma(\pi(n))$. We

can thus write the maximizing permutation π as $\pi(i) = \sum_j \mathbf{1}_{\{j+\sigma(j) \leq i+\sigma(i)\}}$. Hence,

$$\begin{aligned} \sup_{\pi} \sum_i \pi(i) \cdot (i + \sigma(i)) &= \sum_i \sum_j \mathbf{1}_{\{j+\sigma(j) \leq i+\sigma(i)\}} (i + \sigma(i)) \\ &= \frac{1}{2} \sum_{i,j} \max \{i + \sigma(i), j + \sigma(j)\}. \end{aligned}$$

Therefore,

$$\inf_{\pi \in \mathcal{S}_n} \sum_i (i - \pi(i))^2 + (i - \sigma\pi(i))^2 = 4 \sum_i i^2 - \sum_{i,j} \max \{i + \sigma(i), j + \sigma(j)\}. \quad (7.1)$$

Let μ_σ be the empirical distribution associated to σ . If (X, Y) and (X', Y') are two independent random variables with distribution μ_σ then from (7.1) and elementary simplifications we get

$$\begin{aligned} \mathbb{E}[\max \{X + Y, X' + Y'\}] &= \frac{2}{n^2} \sum_{i,j} \max \left\{ \frac{i + \sigma(i)}{n} - 1, \frac{j + \sigma(j)}{n} - 1 \right\} \\ &= \frac{2}{n^3} \left[4 \sum_i i^2 - \inf_{\pi} \sum_i (i - \pi(i))^2 + (i - \sigma(\pi(i)))^2 \right] - 2 \\ &= \frac{2}{n^3} \left[4 \sum_i i^2 - \frac{n^3}{4} W(\mu^{\text{id}^n}, \mu^\sigma)^2 \right] - 2, \end{aligned}$$

where the last equality follows from Lemma 7.2. Since $\sum_i i^2 = n^3/3 + O(n^2)$, the above implies

$$W(\mu^{\text{id}^n}, \mu^\sigma)^2 = \frac{4}{3} - 2\mathbb{E}[\max \{X + Y, X' + Y'\}] + O(1/n).$$

There exists permutations $\sigma^n \in \mathcal{S}_n$ such that μ^{σ^n} converges to \mathbf{P} in the Wasserstein distance by [8, Theorem 1.6] and Lemma 8.1. Therefore, $W(\text{id}, \mathbf{P})^2 = \lim_n \mathbb{E}[W(\mu^{\text{id}^n}, \mu^{\sigma^n})^2]$. Let (X_n, Y_n) be the random variable whose distribution is μ^{σ^n} . Then (X_n, Y_n, X'_n, Y'_n) converges weakly to (X, Y, X', Y') and we conclude from this weak convergence and the formula derived above that

$$\begin{aligned} W(\text{id}, \mathbf{P})^2 &= \lim_{n \rightarrow \infty} \mathbb{E}[W(\mu^{\text{id}^n}, \mu^{\sigma^n})^2] \\ &= \lim_{n \rightarrow \infty} \frac{4}{3} - 2\mathbb{E}[\max \{X_n + Y_n, X'_n + Y'_n\}] + O(1/n) \\ &= \frac{4}{3} - 2\mathbb{E}[\max \{X + Y, X' + Y'\}]. \end{aligned} \quad \square$$

Observe that for a permutation $\mathbf{P} = (X, Y)$ we have $W((X, Y), \text{rev}) = W((X, -Y), \text{id})$. From Lemma 7.3 we conclude that for any permutation $\mathbf{P} = (X, Y)$,

$$W(\text{id}, \mathbf{P})^2 + W(\mathbf{P}, \text{rev})^2 = \frac{8}{3} - 2\mathbb{E}[\max \{X + Y, X' + Y'\} + \max \{X - Y, X' - Y'\}]. \quad (7.2)$$

Proof of Theorem 7.1. Given $\mathbf{P} = (X, Y)$, let $W = \frac{X-Y}{\sqrt{2}}$, $V = \frac{X+Y}{\sqrt{2}}$. Define (W', V') analogously for

the pair (X', Y') . From (7.2) we have

$$W(\mathbf{id}, \mathbf{P})^2 + W(\mathbf{P}, \mathbf{rev})^2 = \frac{8}{3} - 2\sqrt{2} \mathbb{E}[\max\{W, W'\} + \max\{V, V'\}].$$

Suppose Z is an integrable random variable and let Z' be an independent copy of Z . Then

$$\mathbb{E}[\max\{Z, Z'\}] = 2\mathbb{E}[Z \mathbb{P}_{Z'}[Z' < Z]] + \mathbb{E}[Z \mathbb{P}_{Z'}[Z' = Z]].$$

Define $F(z, u) = \mathbb{P}[Z < z] + u\mathbb{P}[Z = z]$ for $z \in \mathbb{R}$ and $0 < u < 1$. The function F is called the “distributional transform” of Z . If U is independent of Z and distributed as Uniform $[0, 1]$ then $F(Z, U)$ is distributed as Uniform $[0, 1]$ as well [16, Proposition 2.1]. Choosing $U \sim \text{Uniform}[0, 1]$ independently of both Z and Z' , and using $\mathbb{P}[Z' < z] = F(z, u) - u\mathbb{P}[Z' = z]$, we get that

$$\mathbb{E}[Z \mathbb{P}[Z' < Z]] = \mathbb{E}[ZF(Z, U)] - \frac{1}{2}\mathbb{E}[Z \mathbb{P}_{Z'}[Z = Z']].$$

In particular, $\mathbb{E}[\max\{Z, Z'\}] = 2\mathbb{E}[ZF(Z, U)]$. Hence,

$$\mathbb{E}[\max\{W, W'\} + \max\{V, V'\}] = 2\mathbb{E}[WF(W, U) + VG(V, U)],$$

where F and G are the distributional transforms of W and V , respectively, and U is independent of both W and V .

Set $\hat{F}(x, u) = 2F(x, u) - 1$ and $\hat{G}(x, u) = 2G(x, u) - 1$. Then $\hat{F}(W, U)$ and $\hat{G}(V, U)$ are both distributed as Uniform $[-1, 1]$. Observe that

$$WF(W, U) + VG(V, U) = \frac{W\hat{F}(W, U)}{2} + \frac{W}{2} + \frac{VG(V, U)}{2} + \frac{V}{2}.$$

We take expectations of this equation and use that $\mathbb{E}[W] = \mathbb{E}[V] = 0$. Then in order to bound the expectation of the r.h.s. we use the inequality $ab \leq \frac{a^2 + b^2}{2}$. This gives that $W\hat{F}(W, U) \leq (W^2 + \hat{F}(W, U)^2)/2$ and $VG(V, U) \leq (V^2 + \hat{G}(V, U)^2)/2$. Since $\mathbb{E}[W^2 + V^2] = \mathbb{E}[X^2 + Y^2] = 2/3$, we conclude that

$$\mathbb{E}[WF(W, U) + VG(V, U)] \leq \frac{1}{4} \mathbb{E}[W^2 + \hat{F}(W, U)^2 + V^2 + \hat{G}(V, U)^2] = \frac{1}{3}.$$

Furthermore, there is equality if and only if $W = \hat{F}(W, U)$ and $V = \hat{G}(V, U)$, which is equivalent to (W, V) being a permuton. As a result,

$$W(\mathbf{id}, \mathbf{P})^2 + W(\mathbf{P}, \mathbf{rev})^2 \geq \frac{8 - 4\sqrt{2}}{3},$$

with equality if and only if $(\frac{X-Y}{\sqrt{2}}, \frac{X+Y}{\sqrt{2}})$ is a permuton. \square

7.2 Limit of sorting: minimal energy paths from identity

We show that there are minimal energy paths from \mathbf{id} to permutons that have a particular *plank property* which we describe below. Roughly speaking, the mass of the permuton on any negatively

sloped plank must be uniformly bounded by the width of the plank. Permutons with bounded Lebesgue density have this property.

The plank property A plank is a region of $[-1, 1]^2$ enclosed between two parallel lines. Specifically, a plank of slope θ , where $-\pi/2 \leq \theta < \pi/2$, is a region of the form

$$R(\theta, a, b) = \left\{ (x, y) \in [-1, 1]^2 : a \leq \cos\left(\theta + \frac{\pi}{2}\right)x + \sin\left(\theta + \frac{\pi}{2}\right)y \leq b \right\},$$

for some $a \leq b$. The width of this plank is $b - a$. We say the plank is negatively sloped if $\theta \leq 0$ and positively sloped if $\theta \geq 0$. A permuton $\mathbf{P} \sim (X, Y)$ has the plank property if there is a constant C such that for every negatively sloped plank R of width w , $\mathbb{P}[(X, Y) \in R] \leq Cw$, or if the same inequality holds for every positively sloped plank.

Theorem 7.4 (Minimal energy paths). *Suppose a permuton \mathbf{P} has the plank property described above. Then there exists a minimal energy path in \mathcal{P} from \mathbf{id} to \mathbf{P} of finite energy.*

The key ingredient, stated in Lemma 7.5 below, is showing that there is a finite energy path from \mathbf{id} to \mathbf{P} . We expect this to hold for all permutons. The remaining argument follows from compactness.

Proof. The set of finite energy paths in \mathcal{P} from \mathbf{id} to \mathbf{P} is non-empty by Lemma 7.5. Choose a sequence of such paths γ_n with $\mathcal{E}[\gamma_n] \searrow \ell$, where $\ell = \inf \mathcal{E}[\gamma]$, the infimum being over all paths γ from \mathbf{id} to \mathbf{P} . Observe that

$$W(\gamma_n(s), \gamma_n(t)) \leq \sqrt{\mathcal{E}[\gamma_n]} |t - s|^{1/2} \leq \sqrt{\mathcal{E}[\gamma_1]} |t - s|^{1/2}.$$

Now consider the space of paths in \mathcal{P} in the uniform metric: $d(\gamma, \gamma') = \sup_{t \in [0, 1]} W(\gamma(t), \gamma'(t))$. As \mathcal{P} is compact in the Wasserstein metric, the equicontinuity estimate above and Arzela-Ascoli Theorem imply that there is a subsequence γ_{n_i} converging in the metric d to a limit path γ_∞ . Clearly, $\gamma_\infty(0) = \mathbf{id}$ and $\gamma_\infty(1) = \mathbf{P}$. As $W(\gamma_{n_i}(t), \gamma_{n_i}(s)) \rightarrow W(\gamma_\infty(t), \gamma_\infty(s))$, we deduce that for any finite partition Π of $[0, 1]$,

$$\mathcal{E}[\gamma_\infty, \Pi] = \lim_{i \rightarrow \infty} \mathcal{E}[\gamma_{n_i}, \Pi] \leq \lim_{i \rightarrow \infty} \mathcal{E}[\gamma_{n_i}] = \ell.$$

Taking the supremum over all finite partitions gives $\mathcal{E}[\gamma_\infty] \leq \ell$, and thus, $\mathcal{E}[\gamma_\infty] = \ell$. \square

Lemma 7.5. *Let $\mathbf{P} \in \mathcal{P}$ satisfy the plank property. Then there is a path in \mathcal{P} from \mathbf{id} to \mathbf{P} that is Lipschitz, and hence, has finite energy.*

Proof. Let us first consider the case that \mathbf{P} has the plank property for all negatively sloped planks. Let $(X, Y) \sim \mathbf{P}$ and define

$$Z(t) = \cos(\pi t/2)X + \sin(\pi t/2)Y \text{ for } 0 \leq t \leq 1.$$

The plank property implies that $\mathbb{P}[a \leq Z(t) \leq b] \leq C \cdot (b - a)$ for all $a < b$ because the event $\{a \leq Z(t) \leq b\} = \{(X, Y) \in R((t-1)\pi/2, a, b)\}$. In particular, $Z(t)$ has no atoms. Set $F_t(x) = \mathbb{P}[Z(t) \leq x]$. As $Z(t)$ has no atoms, $X(t) := 2F_t(Z(t)) - 1$ is distributed uniformly on $[-1, 1]$. Hence, $t \rightarrow (X, X(t))$ is a \mathcal{P} -valued stochastic process that satisfies $(X, X(0)) \sim \mathbf{id}$ and $(X, X(1)) \sim \mathbf{P}$. Below we bound $|F_t(b) - F_s(a)|^2$ in order to get a bound on $\mathbb{E}[|X(t) - X(s)|^2]$.

In the following calculations $C = C_{\mathbf{P}}$ will denote a constant whose value may change from line to line. Note for $a \leq b$,

$$F_t(b) - F_t(a) = \mathbb{P}[(X, Y) \in R((t-1)\pi/2, a, b)] \leq C(b-a).$$

Also, for $s \leq t$ we have

$$|F_t(a) - F_s(a)| = \left| \mathbb{P} \left[(X, Y) \in R \left(\frac{(t-1)\pi}{2}, -\sqrt{2}, a \right) \setminus R \left(\frac{(s-1)\pi}{2}, -\sqrt{2}, a \right) \right] \right. \\ \left. - \mathbb{P} \left[(X, Y) \in R \left(\frac{(s-1)\pi}{2}, -\sqrt{2}, a \right) \setminus R \left(\frac{(t-1)\pi}{2}, -\sqrt{2}, a \right) \right] \right|.$$

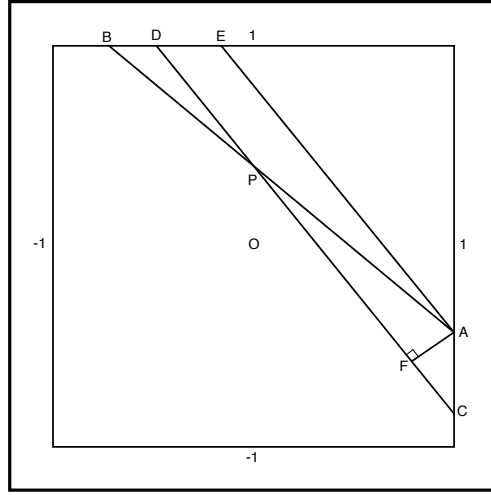


Figure 4: Points (x, y) on the segment AB satisfy $\cos(\pi t/2)x + \sin(\pi t/2)y = a$ and points (x, y) on the segment CD satisfy $\cos(\pi s/2)x + \sin(\pi s/2)y = a$. The region R_1 is the intersection of the convex cone generated by the vectors \vec{PA} and \vec{PC} with $[-1, 1]^2$, and centered at P . The region R_2 is analogously formed by the cone generated by vectors \vec{PB} and \vec{PD} . The angles $\angle APC = \angle BPD = \pi(t-s)/2$. Consider the region R_1 . It is enclosed by either a triangle, e.g., APC as in the figure or a quadrilateral with vertices A, P, C and the point $(1, -1)$. The region is then contained in the plank enclosed by the line segments AE and CD . The width of this plank is $|AF| = |AP| \sin(\angle APC) \leq \sqrt{2} \sin(\pi(t-s)/2) \leq 3(t-s)$.

The regions $R_1 := R((t-1)\pi/2, -\sqrt{2}, a) \setminus R((s-1)\pi/2, -\sqrt{2}, a)$ or $R_2 := R((s-1)\pi/2, -\sqrt{2}, a) \setminus R((t-1)\pi/2, -\sqrt{2}, a)$ are each contained in a negatively sloped plank of width at most $3(t-s)$. See Figure 4 for an illustration and details. Therefore, $|F_t(a) - F_s(a)| \leq C(t-s)$. Now, as $(F_t(b) - F_s(a))^2 \leq 2(F_t(b) - F_t(a))^2 + 2(F_t(a) - F_s(a))^2$, we deduce that

$$\mathbb{E}[(X(t) - X(s))^2] = 4 \mathbb{E}[(F_t(Z(t)) - F_s(Z(s)))^2] \leq C \mathbb{E}[(Z(t) - Z(s))^2] + C(t-s)^2.$$

Observe that

$$\mathbb{E}[(Z(t) - Z(s))^2] = \frac{4}{3} \sin\left(\frac{\pi}{4}\right)(t - s)^2 + \mathbb{E}[XY] \cdot \left(\cos\left(\frac{\pi}{2}t\right) - \cos\left(\frac{\pi}{2}s\right)\right) \cdot \left(\sin\left(\frac{\pi}{2}t\right) - \sin\left(\frac{\pi}{2}s\right)\right).$$

As $\max\{|\sin(b-a)|, |\sin(b) - \sin(a)|, |\cos(b) - \cos(a)|\} \leq |b-a|$, and $|\mathbb{E}[XY]| \leq 1/3$, we conclude that $\mathbb{E}[(Z(t) - Z(s))^2] \leq C(t-s)^2$. Thus, $\mathbb{E}[(X(t) - X(s))^2] \leq C(t-s)^2$ for all $0 \leq s \leq t \leq 1$. If $\gamma(t)$ is the distribution of $(X, X(t))$ then γ is a permuton valued path from \mathbf{id} to \mathbf{P} with $W(\gamma(t), \gamma(s)) \leq \mathbb{E}[(X(t) - X(s))^2]^{1/2} \leq C|t-s|$.

Now if \mathbf{P} has the plank property for all positively sloped planks then we first go from the identity permuton to the reverse permuton along the Archimedean path, and then use the obvious variant of the construction above to find a finite energy path from the reverse permuton to \mathbf{P} . \square

Remark: The above construction is the permuton process limit of so-called stretchable sorting networks. For a target permuton \mathbf{P} the stretchable sorting network of \mathcal{S}_n is defined by sampling n times from \mathbf{P} , rotating the points by angles $\theta \in [0, \pi/2]$, and then taking the ordering of the x -coordinates of each rotated point set. This shows that the Archimedean path can be realized as a limit of some sorting network.

Acknowledgements

M. Rahman's research was partially supported by an NSERC PDF award. B. Virág's research was partially supported by Marie Skłodowska-Curie actions "Spectra" and the NSERC Discovery Accelerator Supplements Program. M. Vizer was supported by the Hungarian National Research, Development and Innovation Office – NKFIH under the grant SNN 116095.

8 Appendix

Proof of Lemma 2.1

Proof. Note that $m^\delta(Y)$ is non increasing in δ and converges to zero as $\delta \rightarrow 0$ almost surely since $t \rightarrow Y(t)$ is uniformly continuous. Also, $m^\delta(Y) \leq 2$, and thus, $\mathbb{E}[m^\delta(Y)] \rightarrow 0$ as $\delta \rightarrow 0$ by the bounded convergence Theorem. For the second claim observe that $||Y(t) - Y(s)| - |\hat{Y}(t) - \hat{Y}(s)|| \leq 2||Y - \hat{Y}||_\infty$ by the triangle inequality. Therefore, $m^\delta(Y) \leq m^\delta(\hat{Y}) + 2||Y - \hat{Y}||_\infty$ and vice-versa, which implies the claim. Finally, for the third claim notice that

$$|\text{Lin}(n, Y)(t) - Y(t)| = \sum_{i=1}^n |Y(\frac{i}{n}) - Y(t) + n(Y(\frac{i-1}{n}) - Y(\frac{i}{n}))(t - \frac{i}{n})| \mathbf{1}_{[\frac{i-1}{n}, \frac{i}{n}]}(t) \leq 2m^{1/n}(Y).$$

The claim now follows from the assertion of the first claim. \square

Proof of Lemma 3.1

Proof. This follows from the Cauchy-Schwarz inequality. Suppose that $\Pi = \{0 = t_0 < t_1 \dots < t_n = 1\}$. As Π' is a refinement of Π it contains points between the t_i . Suppose the points of Π' are indexed as

$t_{i,j}$ for $0 \leq i \leq n$ and $0 \leq j \leq k_i$ such that $t_i = t_{i,0} < t_{i,1} < \dots < t_{i,k_i} = t_{i+1,0} = t_{i+1}$. From the triangle inequality, $d(\gamma(t_i), \gamma(t_{i-1})) \leq \sum_{j=1}^{k_{i-1}} d(\gamma(t_{i-1,j}), \gamma(t_{i-1,j-1}))$.

So from the Cauchy-Schwarz inequality we deduce that

$$\begin{aligned} \sum_{j=1}^{k_{i-1}} d(\gamma(t_{i-1,j}), \gamma(t_{i-1,j-1})) &= \sum_{j=1}^{k_{i-1}} \sqrt{t_{i-1,j} - t_{i-1,j-1}} \frac{d(\gamma(t_{i-1,j}), \gamma(t_{i-1,j-1}))}{\sqrt{t_{i-1,j} - t_{i-1,j-1}}} \\ &\leq \left[\sum_{j=1}^{k_{i-1}} (t_{i-1,j} - t_{i-1,j-1}) \right]^{1/2} \left[\sum_{j=1}^{k_{i-1}} \frac{d(\gamma(t_{i-1,j}), \gamma(t_{i-1,j-1}))^2}{t_{i-1,j} - t_{i-1,j-1}} \right]^{1/2} \\ &= \sqrt{t_i - t_{i-1}} \left[\sum_{j=1}^{k_{i-1}} \frac{d(\gamma(t_{i-1,j}), \gamma(t_{i-1,j-1}))^2}{t_{i-1,j} - t_{i-1,j-1}} \right]^{1/2}. \end{aligned}$$

We conclude that

$$\frac{d(\gamma(t_i), \gamma(t_{i-1}))^2}{t_i - t_{i-1}} \leq \sum_{j=1}^{k_{i-1}} \frac{d(\gamma(t_{i-1,j}), \gamma(t_{i-1,j-1}))^2}{t_{i-1,j} - t_{i-1,j-1}}.$$

Summing the inequality above over i we conclude that $\mathcal{E}[\gamma, \Pi] \leq \mathcal{E}[\gamma, \Pi']$. \square

Proof of Lemma 5.1

Proof. Suppose that γ is a path on the unit sphere of a Hilbert space H from the vector $\gamma(0)$ to its antipode $-\gamma(0)$. For unit vectors a and b it is easily seen that $\|a - b\|^2 = 4\sin(\theta/2)^2$ where $\theta = \arccos(\langle a, b \rangle)$ is the unique angle between a and b in the interval $[0, \pi]$. Let $\theta(s, t)$ denote the angle between $\gamma(s)$ and $\gamma(t)$. Using the inequality $x - (x^3/6) \leq \sin(x) \leq x$ for $0 \leq x \leq \pi$, we see that for any partition $\Pi = \{0 = t_0 < \dots < t_n = 1\}$,

$$(1 - \delta(\Pi))^2 \sum_{i=1}^n \frac{\theta(t_{i-1}, t_i)^2}{t_{i-1} - t_i} \leq \mathcal{E}[\gamma, \Pi] \leq \sum_{i=1}^n \frac{\theta(t_{i-1}, t_i)^2}{t_{i-1} - t_i},$$

where $\delta(\Pi) = \max_i \{\theta(t_{i-1}, t_i)^2\}/24$. Note that $\theta(s, t)$ is continuous as γ is continuous, and in particular, $\delta(\Pi) \rightarrow 0$ as the mesh size $\Delta(\Pi) \rightarrow 0$. This implies that $\mathcal{E}[\gamma] = \sup_{\Pi} \sum_i \frac{\theta(t_{i-1}, t_i)^2}{t_{i-1} - t_i}$.

If a, b and c are unit vectors in H then the corresponding angles between them satisfy $\theta(a, b) + \theta(b, c) \geq \theta(a, c)$. This elementary fact can be deduced by considering unit vectors in \mathbb{R}^3 since a, b and c lie in a 3-dimensional subspace (see, for example, [13, chapter 5]). In particular,

$$\sum_i \theta(t_{i-1}, t_i) \geq \theta(0, 1) = \pi$$

because $\gamma(0)$ and $\gamma(1)$ are antipodal. From the Cauchy-Schwarz inequality we conclude that

$$\pi^2 \leq \sum_i (\sqrt{t_{i-1} - t_i})^2 \cdot \sum_i \frac{\theta(t_{i-1}, t_i)^2}{(\sqrt{t_{i-1} - t_i})^2} = \sum_i \frac{\theta(t_{i-1}, t_i)^2}{t_{i-1} - t_i},$$

with equality only if $\theta(t_{i-1}, t_i) = \pi(t_{i-1} - t_i)$. We may now conclude that $\mathcal{E}[\gamma] \geq \pi^2$, and moreover,

if there is equality then $\theta(s, t) = \pi|s - t|$.

Now suppose that $\mathcal{E}[\gamma] = \pi^2$. We show that $\gamma(t) = \cos(\pi t)\gamma(0) + \sin(\pi t)\gamma(1/2)$. From the fact that $\theta(s, t) = \pi|t - s|$ we see that $\theta(0, t) = \pi t$ and we may write

$$\gamma(t) = \cos(\pi t)\gamma(0) + \sin(\pi t)x(t),$$

where $x(t)$ is a unit vector that is orthogonal to $\gamma(0)$. Noting that $x(1/2) = \gamma(1/2)$ we have that $\langle \gamma(t), \gamma(1/2) \rangle = \sin(\pi t) \langle x(t), x(1/2) \rangle$. Now $\langle \gamma(t), \gamma(1/2) \rangle = \cos(\theta(t, 1/2))$, which equals $\sin(\pi t)$ because $\theta(t, 1/2) = \pi|t - \frac{1}{2}|$. Therefore, $\langle x(t), x(1/2) \rangle = 1$ for $0 < t < 1$, which implies that $x(t) = x(1/2)$ as both of these are unit vectors. Consequently, $\gamma(t) = \cos(\pi t)\gamma(0) + \sin(\pi t)\gamma(1/2)$. \square

Proof of Lemma 7.2

Proof. Couplings between μ_σ and μ_τ are supported on the points $\left(\frac{2i}{n} - 1, \frac{2\sigma(i)}{n} - 1, \frac{2j}{n} - 1, \frac{2\tau(j)}{n} - 1\right)$ for $1 \leq i, j \leq n$. This is because μ_σ is supported on the points $\left(\frac{2i}{n} - 1, \frac{2\sigma(i)}{n} - 1\right)$ and similarly for μ_τ . Thus, a coupling (V, W) between μ_σ and μ_τ is described by the array of numbers $[\alpha_{i,j}]_{1 \leq i, j \leq n}$ such that

$$\alpha_{i,j} = \mathbb{P} \left[V = \left(\frac{2i}{n} - 1, \frac{2\sigma(i)}{n} - 1 \right), W = \left(\frac{2j}{n} - 1, \frac{2\tau(j)}{n} - 1 \right) \right]. \quad (8.1)$$

The constraints $V \sim \mu_\sigma$ and $W \sim \mu_\tau$ is equivalent to

$$\sum_{i=1}^n \alpha_{i,j} = \frac{1}{n} \text{ for } 1 \leq j \leq n \text{ and } \sum_{j=1}^n \alpha_{i,j} = \frac{1}{n} \text{ for } 1 \leq i \leq n.$$

Therefore, the matrix $M = [n\alpha_{i,j}]$ is doubly stochastic. Denoting $M = [m_{i,j}]$ we get that

$$\mathbb{E} [||V - W||^2] = \frac{4}{n^3} \sum_{i,j} m_{i,j} [(i - j)^2 + (\sigma(i) - \tau(j))^2]. \quad (8.2)$$

Conversely, any doubly stochastic matrix $M = [m_{i,j}]$ gives a coupling (V, W) between μ_σ and μ_τ by defining $\alpha_{i,j} = m_{i,j}/n$, and then using (8.1) to define the joint distribution of (V, W) . The corresponding value of $\mathbb{E} [||V - W||^2]$ is given by (8.2).

Let \mathcal{B}_n be the set of all $n \times n$ doubly stochastic matrices. This is a compact, convex subset of \mathbb{R}^{n^2} . The Birkhoff-von Neumann Theorem states that the extreme points of \mathcal{B}_n are the permutation matrices P_π , for $\pi \in \mathcal{S}_n$, defined by $P_\pi(i, j) = \mathbf{1}_{\{\pi(i)=j\}}$. For a matrix $M = [m_{i,j}] \in \mathcal{B}_n$,

$$\sum_{i,j} m_{i,j} [(i - j)^2 + (\sigma(i) - \tau(j))^2] = 4 \sum_i i^2 - 2 \sum_{i,j} m_{i,j} [ij + \sigma(i)\tau(j)].$$

Consequently, we may deduce from the definition of the Wasserstein distance and (8.2) that

$$W(\mu_\sigma, \mu_\tau)^2 = \frac{4}{n^3} \left[4 \sum_i i^2 - 2 \sup_{M \in \mathcal{B}_n} \sum_{i,j} m_{i,j} [ij + \sigma(i)\tau(j)] \right].$$

The map $M \rightarrow \sum_{i,j} m_{i,j} (ij + \sigma(i)\tau(j))$ is linear in M . Hence, it is maximized at one of the extreme points of \mathcal{B}_n . For a permutation P_π , we have that $\sum_{i,j} m_{i,j} [(i - j)^2 + (\sigma(i) - \tau(j))^2] =$

$\sum_i (i - \pi(i))^2 + (\sigma(i) - \tau(\pi(i)))^2$. As a result, we conclude from (8.2) that

$$W(\mu_\sigma, \mu_\tau)^2 = \frac{4}{n^3} \left[\inf_{\pi \in \mathcal{S}_n} \sum_i (i - \pi(i))^2 + (\sigma(i) - \tau(\pi(i)))^2 \right].$$

□

Lemma 8.1. *Let (K, d) be a compact metric space. Let ν_n be a sequence of Borel probability measures on K . Then ν_n converges weakly to a measure ν if and only if $W(\nu_n, \nu) \rightarrow 0$.*

Proof. Suppose that $\nu_n \rightarrow \nu$ weakly. Skorokhod's representation Theorem [9, Theorem 3.30] states that there exists random variables V'_n and V' defined on a common probability space such that $V'_n \sim \nu_n$, $V' \sim \nu$, and $V'_n \rightarrow V'$ pointwise almost surely. But $W(\nu_n, \nu)^2 \leq \mathbb{E}[d(V'_n, V')^2]$. Since $d(V'_n, V') \rightarrow 0$ almost surely and $d(V'_n, V') \leq \text{diam}(\hat{K}) < \infty$, the bounded convergence Theorem implies that $W(\nu_n, \nu)^2 \rightarrow 0$.

Conversely, suppose that $W(\nu_n, \nu) \rightarrow 0$. There exists a coupling (V_n, V) of ν_n and ν such that $\mathbb{E}[d(V'_n, V')^2] \leq 2W(\nu_n, \nu)^2$ for every n . Then $\mathbb{P}[d(V'_n, V') > \delta] \leq 2W(\nu_n, \nu)^2/\delta^2$ by Markov's inequality. Let $f : K \rightarrow \mathbb{R}$ be continuous. Given $\epsilon > 0$ there exists a δ such that $|f(u) - f(v)| < \epsilon$ if $d(u, v) < \delta$ due to uniform continuity. As such, $|\int f \nu_n(dx) - \int f \nu(dx)| = |\mathbb{E}[f(V'_n) - f(V)]| \leq \epsilon + 2\|f\|_\infty \mathbb{P}[d(V'_n, V') > \delta] \leq \epsilon + 4\frac{\|f\|_\infty W(\nu_n, \nu)^2}{\delta^2}$. As ϵ was arbitrary it follows that $\lim_{n \rightarrow \infty} |\int f \nu_n(dx) - \int f \nu(dx)| = 0$ for every continuous f , as required. □

Lemma 8.2. *Let ν, ν' be Borel probability measures on a compact metric space (K, d) . There exists a coupling (V, W) of ν with ν' such that $W(\nu, \nu') = \mathbb{E}[d(V, W)^2]^{1/2}$.*

Proof. Consider couplings $(V_n, W_n) \in K^2$ such that $V_n \sim \nu$, $W_n \sim \nu'$ and $\mathbb{E}[d(V_n, W_n)^2] \rightarrow W(\nu, \nu')^2$. K^2 is a compact metric space in the product topology and so $\mathcal{M}(K^2)$ is compact the weak topology (Prokhorov's Theorem). Hence, we can find a subsequence (V_{n_i}, W_{n_i}) that converges weakly to some (V, W) . As the distance function is continuous and K is compact we conclude that $\mathbb{E}[d(V, W)^2] = \lim_{i \rightarrow \infty} \mathbb{E}[d(V_{n_i}, W_{n_i})^2] = W(\nu, \nu')^2$. □

Lemma 8.3. *Let K be a complete metric space and $S \subset [0, 1]$ a countable dense set. Suppose $f : S \rightarrow K$ has modulus of continuity m on S , i.e., $d(f(t), f(s)) \leq m(|t - s|)$ for $s, t \in S$. Then f has an extension to $[0, 1]$ with modulus of continuity m .*

Proof. For $t \in [0, 1]$ let t_n be any sequence in S converging to t . Then $f(t_n)$ is a Cauchy sequence in K since $d(f(t_n), f(t_m)) \leq m(|t_n - t_m|) \rightarrow 0$ as $n, m \rightarrow \infty$. Let $f(t) \in K$ denote the limit, and note that this does not depend on the approximating sequence t_n due to f having modulus of continuity m on S . In this manner f extends to $[0, 1]$. Also, if $s, t \in [0, 1]$ and $s_n \rightarrow s, t_n \rightarrow t$ then $d(f(t), f(s)) = \lim_n d(f(t_n), f(s_n)) \leq \lim_n m(|t_n - s_n|) = m(|t - s|)$. □

References

- [1] O. Angel, V. Gorin, and A. E. Holroyd. A pattern theorem for random sorting networks. *Electronic Journal of Probability*, 17(99):1–16, 2012.
- [2] O. Angel and A. E. Holroyd. Random subnetworks of random sorting networks. *Electronic Journal of Combinatorics*, 17(1):paper 23, 2010.

- [3] O. Angel, A. E. Holroyd, D. Romik, and B. Virág. Random sorting networks. *Advances in Mathematics*, 215(2):839–868, 2007.
- [4] B. B. Bhattacharya and S. Mukherjee. Degree sequence of random permutation graphs. *Annals of Applied Probability*, to appear.
- [5] R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2004.
- [6] P. Edelman and C. Greene. Balanced tableaux. *Advances in Mathematics*, 63(1):42–99, 1987.
- [7] R. Glebov, A. Grzesik, T. Klimošová, and D. Král. Finitely forcible graphons and permutons. *Journal of Combinatorial Theory, Series B.*, 110:112–135, 2015.
- [8] C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Rath, and R. M. Sampaio. Limits of permutation sequences. *Journal of Combinatorial Theory, Series B.*, 102(1):93–113, 2013.
- [9] O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag New York, 2nd edition, 2002.
- [10] R. Kenyon, D. Král, C. Radin, and P. Winkler. Permutations with fixed pattern densities. *preprint*, 2015, arXiv:1506.02340.
- [11] M. Kotowski. *Limits of random permuton processes and large deviations of the interchange process*. PhD thesis, University of Toronto, 2016.
- [12] D. Král’ and O. Pikhurko. Quasirandom permutations are characterized by 4-point densities. *Geometric and Functional Analysis*, 23:570–579, 2013.
- [13] J. M. Lee. *Riemannian Manifolds*. Springer-Verlag New York, 1997.
- [14] S. Mukherjee. Fixed points and cycle structure of random permutations. *Electronic Journal of Probability*, 21:1–18, 2016.
- [15] M. Rahman and B. Virág. Brownian motion as limit of the interchange process. *preprint*, 2016.
- [16] L. Rüschendorf. On the distributional transform, Sklar’s theorem and the empirical copula process. *Journal of Statistical Planning and Inference*, 139(11):3921–3927, 2009.
- [17] R. P. Stanley. On the number of reduced decompositions of elements of coxeter groups. *European Journal of Combinatorics*, 5(4):359–372, 1984.
- [18] S. Starr. Thermodynamic limit for the mallows model on S_n . *Journal of Mathematical Physics*, 50, 2009.

Mustazee Rahman, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139

E-mail: mustazee@mit.edu Web: www.math.mit.edu/~mustazee

Bálint Virág, Department of Mathematics, University of Toronto, Toronto, ON Canada M5S 2E4.
and MTA Alfred Rényi Institute of Mathematics, Budapest, Hungary.

E-mail: balint@math.toronto.edu Web: www.math.toronto.edu/balint

Máté Vizer, MTA Alfred Rényi Institute of Mathematics, Budapest, Hungary.

E-mail: vizermate@gmail.com